

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
3 January 2003 (03.01.2003)

PCT

(10) International Publication Number  
**WO 03/000864 A2**

(51) International Patent Classification<sup>7</sup>: **C12N**

(21) International Application Number: **PCT/US02/21179**

(22) International Filing Date: **20 June 2002 (20.06.2002)**

(25) Filing Language: **English**

(26) Publication Language: **English**

(30) Priority Data:

60/300,518	22 June 2001 (22.06.2001)	US
60/301,787	29 June 2001 (29.06.2001)	US
60/301,792	29 June 2001 (29.06.2001)	US
60/301,892	29 June 2001 (29.06.2001)	US
60/301,893	29 June 2001 (29.06.2001)	US
60/303,405	6 July 2001 (06.07.2001)	US
60/303,442	6 July 2001 (06.07.2001)	US
60/364,438	15 March 2002 (15.03.2002)	US

(71) Applicant (*for all designated States except US*): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **HAFALIA, April, J. A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **WARREN, Bridget, A.** [US/US]; 2250 Homestead Court #2, Los Altos, CA 94024 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue #71, Palo Alto, CA 94303 (US). **ARVIZU, Chandra, S.** [US/US]; 1706 Morocco Drive, San Jose, CA 95125 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **HONCHELL, Cynthia, D.** [US/US]; 400 Laurel Street #203, San Carlos, CA 94070 (US). **LEE, Ernestine, A.** [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **FORSYTHE, Ian, J.** [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **YANG, Junming** [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **SANJANWALA, Madhusudan, M.** [US/US]; 210 Silvia Court, Los Altos, CA 94024 (US). **BAUGHN, Mariah, R.**

[US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street A #205, San Leandro, CA 94577 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **LAL, Preeti, G.** [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **BECHA, Shanya, D.** [US/US]; 21062 Gary Drive #117, Castro Valley, CA 94546 (US). **LEE, Soo, Yeun** [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **LUO, Wen** [CN/US]; 181 Cymbidium Circle, South San Francisco, CA 94080 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **ZEBARJADIAN, Yeganeh** [IR/US]; 830 Junipero Serra Boulevard, San Francisco, CA 94127 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US).

(74) Agents: **HAMLET-COX, Diana et al.**; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).**Published:**— *without international search report and to be republished upon receipt of that report**For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*(54) Title: **NUCLEIC ACID-ASSOCIATED PROTEINS**

(57) Abstract: Various embodiments of the invention provide human nucleic acid-associated proteins (NAAP) and polynucleotides which identify and encode NAAP. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of NAAP.

**WO 03/000864 A2**

## NUCLEIC ACID-ASSOCIATED PROTEINS

### TECHNICAL FIELD

The invention relates to novel nucleic acids, nucleic acid-associated proteins encoded by these  
5 nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and  
prevention of cell proliferative, neurological, reproductive, developmental, autoimmune/inflammatory,  
and DNA repair disorders, and infections. The invention also relates to the assessment of the effects  
of exogenous compounds on the expression of nucleic acids and nucleic acid-associated proteins.

### BACKGROUND OF THE INVENTION

Multicellular organisms are comprised of diverse cell types that differ dramatically both in  
structure and function. The identity of a cell is determined by its characteristic pattern of gene  
expression, and different cell types express overlapping but distinctive sets of genes throughout  
development. Spatial and temporal regulation of gene expression is critical for the control of cell  
15 proliferation, cell differentiation, apoptosis, and other processes that contribute to organismal  
development. Furthermore, gene expression is regulated in response to extracellular signals that  
mediate cell-cell communication and coordinate the activities of different cell types. Appropriate gene  
regulation also ensures that cells function efficiently by expressing only those genes whose functions  
are required at a given time.

20 The cell nucleus contains all of the genetic information of the cell in the form of DNA, and the  
components and machinery necessary for replication of DNA and for transcription of DNA in  
RNA. (See Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing Inc. New York  
NY, pp. 335-399.) DNA is organized into compact structures in the nucleus by interactions with  
various DNA-binding proteins such as histones and non-histone chromosomal proteins.

25 DNA-specific nucleases, DNases, partially degrade these compacted structures prior to DNA  
replication or transcription. DNA replication takes place with the aid of DNA helicases which unwind  
the double-stranded DNA helix, and DNA polymerases that duplicate the separated DNA strands.

#### Transcription Factors

Transcriptional regulatory proteins are essential for the control of gene expression. Some of  
30 these proteins function as transcription factors that initiate, activate, repress, or terminate gene  
transcription. Transcription factors generally bind to the promoter, enhancer, and upstream regulatory  
regions of a gene in a sequence-specific manner, although some factors bind regulatory elements  
within or downstream of a gene coding region. Transcription factors may bind to a specific region of

DNA singly or as a complex with other accessory factors (reviewed in Lewin, B. (1990) Genes IV, Oxford University Press, New York NY, and Cell Press, Cambridge MA, pp. 554-570).

The double helix structure and repeated sequences of DNA create topological and chemical features which can be recognized by transcription factors. These features are hydrogen bond donor  
5 and acceptor groups, hydrophobic patches, major and minor grooves, and regular, repeated stretches of sequence which induce distinct bends in the helix. Typically, transcription factors recognize specific DNA sequence motifs of about 20 nucleotides in length. Multiple, adjacent transcription factor-binding motifs may be required for gene regulation.

Many transcription factors incorporate DNA-binding structural motifs which comprise either a  
10 helices or  $\beta$  sheets that bind to the major groove of DNA. Four well-characterized structural motifs are helix-turn-helix, zinc finger, leucine zipper, and helix-loop-helix. Proteins containing these motifs may act alone as monomers, or they may form homo- or heterodimers that interact with DNA.

The helix-turn-helix motif consists of two  $\alpha$  helices connected at a fixed angle by a short chain of amino acids. One of the helices binds to the major groove. Helix-turn-helix motifs are  
15 exemplified by the homeobox motif which is present in homeodomain proteins. These proteins are critical for specifying the anterior-posterior body axis during development and are conserved throughout the animal kingdom. The Antennapedia and Ultrabithorax proteins of *Drosophila melanogaster* are prototypical homeodomain proteins (Pabo, C.O. and R.T. Sauer (1992) *Annu. Rev. Biochem.* 61:1053-1095).

20 Homeobox genes are a family of highly conserved regulatory genes that encode transcription factors. They are essential during embryonic development. They are important in limb formation and reproductive tract development. They function in uterine receptivity and implantation in mice and probably serve a similar role in humans (Daftary, G. S. and Taylor, H. S. (2000) *Semin. Reprod. Med.* 18:311-320). Homeobox gene mutations play a role in susceptibility to autism (Ingram, J. L. et al.  
25 (2000) *Teratology* 62:393-405) and are implicated in human diseases, such as diabetes to cancer (Cillo, C. et al. (2001) *J. Cell Physiol.* 188:161-169).

The helix-loop-helix motif (HLH) consists of a short  $\alpha$  helix connected by a loop to a longer  $\alpha$  helix. The loop is flexible and allows the two helices to fold back against each other and to bind to DNA. The protooncogene Myc, a transcription factor that activates genes required for cellular  
30 proliferation, contains a prototypical HLH motif.

The zinc finger motif, which binds zinc ions, generally contains tandem repeats of about 30 amino acids consisting of periodically spaced cysteine and histidine residues. Examples of this sequence pattern, designated C2H2 and C3HC4 ("RING" finger), have been described (Lewin,

*supra*). Zinc finger proteins each contain an  $\alpha$  helix and an antiparallel  $\beta$  sheet whose proximity and conformation are maintained by the zinc ion. Contact with DNA is made by the arginine preceding the  $\alpha$  helix and by the second, third, and sixth residues of the  $\alpha$  helix. Variants of the zinc finger motif include poorly defined cysteine-rich motifs which bind zinc or other metal ions. These motifs may not contain histidine residues and are generally nonrepetitive. The zinc finger motif may be repeated in a tandem array within a protein, such that the  $\alpha$  helix of each zinc finger in the protein makes contact with the major groove of the DNA double helix. This repeated contact between the protein and the DNA produces a strong and specific DNA-protein interaction. The strength and specificity of the interaction can be regulated by the number of zinc finger motifs within the protein. Though originally identified in DNA-binding proteins as regions that interact directly with DNA, zinc fingers occur in a variety of proteins that do not bind DNA (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York NY, pp. 447-451). For example, Galcheva-Gargova et al. (1996; Science 272:1797-1802) have identified zinc finger proteins that interact with various cytokine receptors.

The C2H2-type zinc finger signature motif contains a 28 amino acid sequence, including 2 conserved Cys and 2 conserved His residues in a C-2-C-12-H-3-H type motif. The motif generally occurs in multiple tandem repeats. A cysteine-rich domain including the motif Asp-His-His-Cys (DHHC-CRD) has been identified as a distinct subgroup of zinc finger proteins. The DHHC-CRD region has been implicated in growth and development. One DHHC-CRD mutant shows defective function of Ras, a small membrane-associated GTP-binding protein that regulates cell growth and differentiation, while other DHHC-CRD proteins probably function in pathways not involving Ras (Bartels, D.J. et al. (1999) Mol. Cell Biol. 19:6775-6787).

Zinc-finger transcription factors are often accompanied by modular sequence motifs such as the Kruppel-associated box (KRAB) and the SCAN domain. For example, the hypoalphalipoproteinemia susceptibility gene ZNF202 encodes a SCAN box and a KRAB domain followed by eight C2H2 zinc-finger motifs (Honer, C. et al. (2001) Biochim. Biophys. Acta 1517:441-448). The SCAN domain is a highly conserved, leucine-rich motif of approximately 60 amino acids found at the amino-terminal end of zinc finger transcription factors. SCAN domains are most often linked to C2H2 zinc finger motifs through their carboxyl-terminal end. Biochemical binding studies have established the SCAN domain as a selective hetero- and homotypic oligomerization domain. SCAN domain-mediated protein complexes may function to modulate the biological function of transcription factors (Schumacher, C. et al. (2000) J. Biol. Chem. 275:17173-17179).

The KRAB (Kruppel-associated box) domain is a conserved amino acid sequence spanning

approximately 75 amino acids and is found in almost one-third of the 300 to 700 genes encoding C2H2 zinc fingers. The KRAB domain is found N-terminally with respect to the finger repeats. The KRAB domain is generally encoded by two exons; the KRAB-A region or box is encoded by one exon and the KRAB-B region or box is encoded by a second exon. The function of the KRAB domain is the repression of transcription. Transcription repression is accomplished by recruitment of either the KRAB-associated protein-1, a transcriptional corepressor, or the KRAB-A interacting protein. Proteins containing the KRAB domain are likely to play a regulatory role during development (Williams, A.J. et al. (1999) *Mol. Cell Biol.* 19:8526-8535). A subgroup of highly related human KRAB zinc finger proteins detectable in all human tissues is highly expressed in human T lymphoid cells (Bellefroid, E.J. et al. (1993) *EMBO J.* 12:1363-1374). The ZNF85 KRAB zinc finger gene, a member of the human ZNF91 family, is highly expressed in normal adult testis, in seminomas, and in the NT2/D1 teratocarcinoma cell line (Poncelet, D.A. et al. (1998) *DNA Cell Biol.* 17:931-943).

The C4 motif is found in hormone-regulated proteins. The C4 motif generally includes only 2 repeats. A number of eukaryotic and viral proteins contain a conserved cysteine-rich domain of 40 to 60 residues (called C3HC4 zinc-finger or RING finger) that binds two atoms of zinc, and is probably involved in mediating protein-protein interactions. The 3D "cross-brace" structure of the zinc ligation system is unique to the RING domain. The spacing of the cysteines in such a domain is C-x(2)-C-x(9 to 39)-C-x(1 to 3)-H-x(2 to 3)-C-x(2)-C-x(4 to 48)-C-x(2)-C. The PHD finger is a C4HC3 zinc-finger-like motif found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation.

GATA-type transcription factors contain one or two zinc finger domains which bind specifically to a region of DNA that contains the consecutive nucleotide sequence GATA. NMR studies indicate that the zinc finger comprises two irregular anti-parallel  $\beta$  sheets and an  $\alpha$  helix, followed by a long loop to the C-terminal end of the finger (Ominchinski, J.G. (1993) *Science* 261:438-446). The helix and the loop connecting the two  $\beta$ -sheets contact the major groove of the DNA, while the C-terminal part, which determines the specificity of binding, wraps around into the minor groove.

The LIM motif consists of about 60 amino acid residues and contains seven conserved cysteine residues and a histidine within a consensus sequence (Schmeichel, K.L. and M.C. Beckerle (1994) *Cell* 79:211-219). The LIM family includes transcription factors and cytoskeletal proteins which may be involved in development, differentiation, and cell growth. One example is actin-binding LIM protein, which may play roles in regulation of the cytoskeleton and cellular morphogenesis (Roof, D.J. et al. (1997) *J. Cell Biol.* 138:575-588). The N-terminal domain of actin-binding LIM protein has

four double zinc finger motifs with the LIM consensus sequence. The C-terminal domain of actin-binding LIM protein shows sequence similarity to known actin-binding proteins such as dematin and villin. Actin-binding LIM protein binds to F-actin through its dematin-like C-terminal domain. The LIM domain may mediate protein-protein interactions with other LIM-binding proteins.

5           Myeloid cell development is controlled by tissue-specific transcription factors. Myeloid zinc finger proteins (MZF) include MZF-1 and MZF-2. MZF-1 functions in regulation of the development of neutrophilic granulocytes. A murine homolog MZF-2 is expressed in myeloid cells, particularly in the cells committed to the neutrophilic lineage. MZF-2 is down-regulated by G-CSF and appears to have a unique function in neutrophil development (Murai, K. et al. (1997) *Genes Cells* 2:581-591).

10           The leucine zipper motif comprises a stretch of amino acids rich in leucine which can form an amphipathic  $\alpha$  helix. This structure provides the basis for dimerization of two leucine zipper proteins. The region adjacent to the leucine zipper is usually basic, and upon protein dimerization, is optimally positioned for binding to the major groove. Proteins containing such motifs are generally referred to as bZIP transcription factors. The leucine zipper motif is found in the proto-oncogenes Fos and Jun,  
15           which comprise the heterodimeric transcription factor AP1 involved in cell growth and the determination of cell lineage (Papavassiliou, A.G. (1995) *N. Engl. J. Med.* 332:45-47).

          The helix-loop-helix motif (HLH) consists of a short  $\alpha$  helix connected by a loop to a longer  $\alpha$  helix. The loop is flexible and allows the two helices to fold back against each other and to bind to DNA. The transcription factor Myc contains a prototypical HLH motif.

20           The NF-kappa-B/Rel signature defines a family of eukaryotic transcription factors involved in oncogenesis, embryonic development, differentiation and immune response. Most transcription factors containing the Rel homology domain (RHD) bind as dimers to a consensus DNA sequence motif termed kappa-B. Members of the Rel family share a highly conserved 300 amino acid domain termed the Rel homology domain. The characteristic Rel C-terminal domain is involved in gene activation and  
25           cytoplasmic anchoring functions. Proteins known to contain the RHD domain include vertebrate nuclear factor NF-kappa-B, which is a heterodimer of a DNA-binding subunit and the transcription factor p65, mammalian transcription factor RelB, and vertebrate proto-oncogene c-rel, a protein associated with differentiation and lymphopoiesis (Kabrun, N. and P.J. Enrietto (1994) *Semin. Cancer Biol.* 5:103-112).

30           A DNA binding motif termed ARID (AT-rich interactive domain) distinguishes an evolutionarily conserved family of proteins. The approximately 100-residue ARID sequence is present in a series of proteins strongly implicated in the regulation of cell growth, development, and tissue-specific gene expression. ARID proteins include Bright (a regulator of B-cell-specific gene

expression), dead ringer (involved in development), and MRF-2 (which represses expression from the cytomegalovirus enhancer) (Dallas, P.B. et al. (2000) Mol. Cell. Biol. 20:3137-3146).

The ELM2 (Egl-27 and MTA1 homology 2) domain is found in metastasis-associated protein MTA1 and protein ER1. The *Caenorhabditis elegans* gene *egl-27* is required for embryonic  
5 patterning MTA1, a human gene with elevated expression in metastatic carcinomas, is a component of a protein complex with histone deacetylase and nucleosome remodelling activities (Solari, F. et al. (1999) Development 126:2483-2494). The ELM2 domain is usually found to the N terminus of a myb-like DNA binding domain. ELM2 is also found associated with an ARID DNA.

The Iroquois (Irx) family of genes are found in nematodes, insects and vertebrates. Irx genes  
10 usually occur in one or two genomic clusters of three genes each and encode transcriptional controllers that possess a characteristic homeodomain. The Irx genes function early in development to specify the identity of diverse territories of the body. Later in development in both *Drosophila* and vertebrates, the Irx genes function again to subdivide those territories into smaller domains (reviewed in Cavodeassi, F. et al. (2001) Development 128:2847-2855). For example, mouse and human Irx4  
15 proteins are 83% conserved and their 63-aa homeodomain is more than 93% identical to that of the *Drosophila* Iroquois patterning genes. Irx4 transcripts are predominantly expressed in the cardiac ventricles. The homeobox gene Irx4 mediates ventricular differentiation during cardiac development (Bruneau, B.G. et al. (2000) Dev. Biol. 217:266-77).

Histidine triad (HIT) proteins share residues in distinctive dimeric, 10-stranded half-barrel  
20 structures that form two identical purine nucleotide-binding sites. Hint (histidine triad nucleotide-binding protein)-related proteins, found in all forms of life, and fragile histidine triad (Fhit)-related proteins, found in animals and fungi, represent the two main branches of the HIT superfamily. Fhit homologs bind and cleave diadenosine polyphosphates. Fhit-Ap(n)A complexes appear to function in a proapoptotic tumor suppression pathway in epithelial tissues (Brenner C. et al.  
25 (1999) J. Cell Physiol. 181:179-187).

Most transcription factors contain characteristic DNA binding motifs, and variations on the above motifs and new motifs have been and are currently being characterized (Faisst, S. and S. Meyer (1992) Nucleic Acids Res. 20:3-26). These include the forkhead motif, found in transcription factors involved in development and oncogenesis (Hacker et al. (1995) EMBO J 14:5306-5317), and  
30 the T-box protein T-domain, which forms a novel major and minor groove DNA contact. T-box genes such as *Brachyury* (*T*) are essential for tissue specification in development. (Muller (1997) Nature 389:884-888.)

#### Chromatin Associated Proteins

In the nucleus, DNA is packaged into chromatin, the compact organization of which limits the accessibility of DNA to transcription factors and plays a key role in gene regulation (Lewin, *supra*, pp. 409-410). The compact structure of chromatin is determined and influenced by chromatin-associated proteins such as the histones, the high mobility group (HMG) proteins, and the chromodomain proteins. There are five classes of histones, H1, H2A, H2B, H3, and H4, all of which are highly basic, low molecular weight proteins. The fundamental unit of chromatin, the nucleosome, consists of 200 base pairs of DNA associated with two copies each of H2A, H2B, H3, and H4. H1 links adjacent nucleosomes. HMG proteins are low molecular weight, non-histone proteins that may play a role in unwinding DNA and stabilizing single-stranded DNA. Chromodomain proteins play a key role in the formation of highly compacted heterochromatin, which is transcriptionally silent.

#### **Diseases and Disorders Related to Gene Regulation**

Mutations in transcription factors contribute to oncogenesis. This is likely due to the role of transcription factors in the expression of genes involved in cell proliferation. For example, mutations in transcription factors encoded by proto-oncogenes, such as Fos, Jun, Myc, Rel, and Sp1, may be oncogenic due to increased stimulation of cell proliferation. Conversely, mutations in transcription factors encoded by tumor suppressor genes, such as p53, RB1, and WT1, may be oncogenic due to decreased inhibition of cell proliferation. (Latchman, D. (1995) Gene Regulation: A Eukaryotic Perspective, Chapman and Hall, London, UK, pp 242-255.)

Many neoplastic disorders in humans can be attributed to inappropriate gene expression. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes (Cleary, M.L. (1992) *Cancer Surv.* 15:89-104). The zinc finger-type transcriptional regulator WT1 is a tumor-suppressor protein that is inactivated in children with Wilm's tumor. Deletions of the WT1 gene, or point mutations which destroy the DNA-binding activity of the protein, are associated with development of the pediatric nephroblastoma, Wilms tumor, and Denys-Drash syndrome. (Rauscher, F.J. (1993) *FASEB J.* 7:896-903.) The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, A.G. (1995) *N. Engl. J. Med.* 332:45-47). Chromosomal translocations may also produce chimeric loci that fuse the coding sequence of one gene with the regulatory regions of a second unrelated gene. Such an arrangement likely results in inappropriate gene transcription, potentially contributing to malignancy. In Burkitt's lymphoma, for example, the transcription factor Myc is translocated to the immunoglobulin heavy chain locus, greatly enhancing Myc expression and resulting in rapid cell growth leading to leukemia (Latchman, D.S. (1996) *N. Engl. J. Med.* 334:28-33).

Certain proteins enriched in glutamine are associated with various neurological disorders

including spinocerebellar ataxia, bipolar effective disorder, schizophrenia, and autism. (Margolis, R.L. et al. (1997) Human Genetics 100:114-122.) These proteins contain regions with as many as 15 or more consecutive glutamine residues and may function as transcription factors with a potential role in regulation of neurodevelopment or neuroplasticity.

5 In addition, the immune system responds to infection or trauma by activating a cascade of events that coordinate the progressive selection, amplification, and mobilization of cellular defense mechanisms. A complex and balanced program of gene activation and repression is involved in this process. However, hyperactivity of the immune system as a result of improper or insufficient regulation of gene expression may result in considerable tissue or organ damage. This damage is well-  
10 documented in immunological responses associated with arthritis, allergens, heart attack, stroke, and infections (Isselbacher, K.J. et al. Harrison's Principles of Internal Medicine, 13/e, McGraw Hill, Inc. and Teton Data Systems Software, 1996). In particular, a zinc finger protein termed Staf50 (for Stimulated trans-acting factor of 50 kDa) is a transcriptional regulator and is induced in various cell lines by interferon-I and -II. Staf50 appears to mediate the antiviral activity of interferon by  
15 down-regulating the viral transcription directed by the long terminal repeat promoter region of human immunodeficiency virus type-1 in transfected cells. (Tissot, C. (1995) J. Biol. Chem. 270:14891-14898.) Also, the causative gene for autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) was recently isolated and found to encode a protein with two PHD-type zinc finger motifs (Bjorses, P. et al. (1998) Hum. Mol. Genet. 7:1547-1553).

20 Furthermore, the generation of multicellular organisms is based upon the induction and coordination of cell differentiation at the appropriate stages of development. Central to this process is differential gene expression, which confers the distinct identities of cells and tissues throughout the body. Failure to regulate gene expression during development could result in developmental disorders. Human developmental disorders caused by mutations in zinc finger-type transcriptional regulators  
25 include: urogenital developmental abnormalities associated with WT1; Greig cephalopolysyndactyly, Pallister-Hall syndrome, and postaxial polydactyly type A (GLI3), and Townes-Brocks syndrome, characterized by anal, renal, limb, and ear abnormalities (SALL1) (Engelkamp, D. and V. van Heyningen (1996) Curr. Opin. Genet. Dev. 6:334-342; Kohlhase, J. et al. (1999) Am. J. Hum. Genet. 64:435-445).

30 Human acute leukemias involve reciprocal chromosome translocations that fuse the ALL-1 gene located at chromosome region 11q23 to a series of partner genes positioned on a variety of human chromosomes. The fused genes encode chimeric proteins. The AF17 gene encodes a protein of 1093 amino acids, containing a leucine-zipper dimerization motif located 3' of the fusion point and a

cysteine-rich domain at the N terminus that shows homology to a domain within the protein Br140 (peregrin) (Prasad R. et al. (1994) Proc. Natl. Acad. Sci. USA 91:8107-8111).

Impaired transcriptional regulation may lead to Alzheimer's disease, a progressive neurodegenerative disorder that is characterized by the formation of senile plaques and neurofibrillary tangles containing amyloid beta peptide. These plaques are found in limbic and association cortices of the brain, including hippocampus, temporal cortices, cingulate cortex, amygdala, nucleus basalis and locus caeruleus. Early in Alzheimer's pathology, physiological changes are visible in the cingulate cortex (Minoshima, S. et al. (1997) Ann. Neurol. 42:85-94). In subjects with advanced Alzheimer's disease, accumulating plaques damage the neuronal architecture in limbic areas and eventually cripple the memory process.

## SYNTHESIS OF NUCLEIC ACIDS

### Polymerases

DNA and RNA replication are critical processes for cell replication and function. DNA and RNA replication are mediated by the enzymes DNA and RNA polymerase, respectively, by a "templating" process in which the nucleotide sequence of a DNA or RNA strand is copied by complementary base-pairing into a complementary nucleic acid sequence of either DNA or RNA. However, there are fundamental differences between the two processes.

DNA polymerase catalyzes the stepwise addition of a deoxyribonucleotide to the 3'-OH end of a polynucleotide strand (the primer strand) that is paired to a second (template) strand. The new DNA strand therefore grows in the 5' to 3' direction (Alberts, B. et al. (1994) The Molecular Biology of the Cell, Garland Publishing Inc., New York NY, pp 251-254). The substrates for the polymerization reaction are the corresponding deoxynucleotide triphosphates which must base-pair with the correct nucleotide on the template strand in order to be recognized by the polymerase. Because DNA exists as a double-stranded helix, each of the two strands may serve as a template for the formation of a new complementary strand. Each of the two daughter cells of a dividing cell therefore inherits a new DNA double helix containing one old and one new strand. Thus, DNA is said to be replicated "semiconservatively" by DNA polymerase. In addition to the synthesis of new DNA, DNA polymerase is also involved in the repair of damaged DNA as discussed below under "Ligases."

In contrast to DNA polymerase, RNA polymerase uses a DNA template strand to "transcribe" DNA into RNA using ribonucleotide triphosphates as substrates. Like DNA polymerization, RNA polymerization proceeds in a 5' to 3' direction by addition of a ribonucleoside monophosphate to the 3'-OH end of a growing RNA chain. DNA transcription generates messenger RNAs (mRNA) that carry information for protein synthesis, as well as the transfer, ribosomal, and

other RNAs that have structural or catalytic functions. In eukaryotes, three discrete RNA polymerases synthesize the three different types of RNA (Alberts, *supra*, pp. 367-368). RNA polymerase I makes the large ribosomal RNAs, RNA polymerase II makes the mRNAs that will be translated into proteins, and RNA polymerase III makes a variety of small, stable RNAs, including 5S ribosomal RNA and the transfer RNAs (tRNA). In all cases, RNA synthesis is initiated by binding of the RNA polymerase to a promoter region on the DNA and synthesis begins at a start site within the promoter. Synthesis is completed at a stop (termination) signal in the DNA whereupon both the polymerase and the completed RNA chain are released.

### Ligases

DNA repair is the process by which accidental base changes, such as those produced by oxidative damage, hydrolytic attack, or uncontrolled methylation of DNA, are corrected before replication or transcription of the DNA can occur. Because of the efficiency of the DNA repair process, fewer than one in a thousand accidental base changes causes a mutation (Alberts, *supra*, pp. 245-249). The three steps common to most types of DNA repair are (1) excision of the damaged or altered base or nucleotide by DNA nucleases, (2) insertion of the correct nucleotide in the gap left by the excised nucleotide by DNA polymerase using the complementary strand as the template and, (3) sealing the break left between the inserted nucleotide(s) and the existing DNA strand by DNA ligase. In the last reaction, DNA ligase uses the energy from ATP hydrolysis to activate the 5' end of the broken phosphodiester bond before forming the new bond with the 3'-OH of the DNA strand. In Bloom's syndrome, an inherited human disease, individuals are partially deficient in DNA ligation and consequently have an increased incidence of cancer (Alberts, *supra*, p. 247).

### Nucleases

Nucleases comprise enzymes that hydrolyze both DNA (DNase) and RNA (Rnase). They serve different purposes in nucleic acid metabolism. Nucleases hydrolyze the phosphodiester bonds between adjacent nucleotides either at internal positions (endonucleases) or at the terminal 3' or 5' nucleotide positions (exonucleases). A DNA exonuclease activity in DNA polymerase, for example, serves to remove improperly paired nucleotides attached to the 3'-OH end of the growing DNA strand by the polymerase and thereby serves a "proofreading" function. As mentioned above, DNA endonuclease activity is involved in the excision step of the DNA repair process.

RNases also serve a variety of functions. For example, RNase P is a ribonucleoprotein enzyme which cleaves the 5' end of pre-tRNAs as part of their maturation process. RNase H digests the RNA strand of an RNA/DNA hybrid. Such hybrids occur in cells invaded by retroviruses, and RNase H is an important enzyme in the retroviral replication cycle. Pancreatic RNase secreted by

the pancreas into the intestine hydrolyzes RNA present in ingested foods. RNase activity in serum and cell extracts is elevated in a variety of cancers and infectious diseases (Schein, C.H. (1997) Nat. Biotechnol. 15:529-536). Regulation of RNase activity is being investigated as a means to control tumor angiogenesis, allergic reactions, viral infection and replication, and fungal infections.

## 5 MODIFICATION OF NUCLEIC ACIDS

### DNA Repair

Cells are constantly faced with replication errors and environmental assault (such as ultraviolet irradiation) that can produce DNA damage. Damage to DNA consists of any change that modifies the structure of the molecule. Changes to DNA can be divided into two general classes, single base changes and structural distortions. Single base changes affect the sequence but not the overall structure of the DNA. Since single base changes do not affect transcription or replication, they exert their effect on future generations. Structural distortions affect the structure of the DNA. A single strand nick or removal of a base may prevent a strand from acting as a viable template for synthesis of DNA or RNA. Intrastrand or interstrand covalent linkage between bases, or the addition of a bulky adduct to a base, may distort the structure of the double helix and interfere with transcription and replication. Any damage to DNA can produce a mutation, and the mutation may produce a disorder, such as cancer.

Changes in DNA are recognized by repair systems within the cell. These repair systems act to correct the damage and thus prevent any deleterious effects of a mutational event. Repair systems can be divided into three general types, direct repair, excision repair, and retrieval systems. When the repair systems are eliminated, cells become exceedingly sensitive to environmental mutagens, such as ultraviolet irradiation. Disorders associated with a loss in DNA repair systems often exhibit a high sensitivity to environmental mutagens. Examples of such disorders include xeroderma pigmentosum, Bloom's syndrome, and Werner's syndrome. Xeroderma pigmentosum results in a hypersensitivity to sunlight, especially ultraviolet, and produces skin defects. Bloom's syndrome results in an increased frequency of chromosomal aberrations, including sister chromosome exchanges (Yamagata, K. et al. (1998) Proc. Natl. Acad. Sci. USA 95:8733-8738).

Direct repair involves the reversal or simple removal of the damaged region of DNA. Mismatches involving normal bases are repaired based on certain biases within the repair system. For example, mismatched GT base pairs are frequently caused by deamination of 5-methyl-cytosine to form thymine. Therefore, repair systems convert mismatched GT pairs to GC, instead of AT. Repair also favors the non-methylated strand in hemimethylated DNA, since this strand represents the newly synthesized daughter strand. The recognition of hemimethylated DNA and repair of mismatches on

the non-methylated strand involve the products of the genes *mutH*, *mutL*, *mutS* (which specifically recognizes mismatched base pairs), the helicase encoded by the *uvrD* gene, and the methylase encoded by the *dam* gene. C-5 cytosine-specific DNA methylases are enzymes that specifically methylate the C-5 carbon of cytosines in DNA (Kumar, S. et al. (1994) *Nucleic Acids Res.* 22:1-10).

5       Excision repair is a system in which mispaired or damaged bases are removed from DNA and a new stretch of DNA is synthesized to replace them. In the incision step, the damaged structure is recognized by an endonuclease that cleaves the DNA strand on both sides of the damage. In the excision step, a 5'-3' exonuclease removes a stretch of the damaged DNA strand. In the synthesis step, the resulting single-stranded region serves as a template for a DNA polymerase to synthesize a replacement for the excised sequence. Finally, DNA ligase covalently links the 3' end of the new material to the old material. In mammals, DNA polymerase beta serves as the DNA repair polymerase. Mutations in the human DNA polymerase beta gene are associated with several types of cancer (Bhattacharyya, N. et al. (1999) *DNA Cell Biol.* 18:549-554; Matsuzaki, J. et al. (1996) *Mol. Carcinog.* 15:38-43).

#### 15   Methylases

Methylation of specific nucleotides occurs in both DNA and RNA, and serves different functions in the two macromolecules. Methylation of cytosine residues to form 5-methyl cytosine in DNA occurs specifically in CG sequences which are base-paired with one another in the DNA double-helix. The pattern of methylation is passed from generation to generation during DNA replication by an enzyme called "maintenance methylase" that acts preferentially on those CG sequences that are base-paired with a CG sequence that is already methylated. Such methylation appears to distinguish active from inactive genes by preventing the binding of regulatory proteins that "turn on" the gene, but permitting the binding of proteins that inactivate the gene (Alberts, *supra*, pp. 448-451). In RNA metabolism, "tRNA methylase" produces one of several nucleotide modifications in tRNA that affect the conformation and base-pairing of the molecule and facilitate the recognition of the appropriate mRNA codons by specific tRNAs. The primary methylation pattern is the dimethylation of guanine residues to form N,N-dimethyl guanine.

#### Helicases and Single-stranded Binding Proteins

Helicases are enzymes that destabilize and unwind double helix structures in both DNA and RNA. Since DNA replication occurs more or less simultaneously on both strands, the two strands must first separate to generate a replication "fork" for DNA polymerase to act on. Two types of replication proteins contribute to this process, DNA helicases and single-stranded binding proteins. DNA helicases hydrolyze ATP and use the energy of hydrolysis to separate the DNA strands.

Single-stranded binding proteins (SSBs) then bind to the exposed DNA strands, without covering the bases, thereby temporarily stabilizing them for templating by the DNA polymerase (Alberts, *supra*, pp. 255-256).

RNA helicases also alter and regulate RNA conformation and secondary structure. Like the  
5 DNA helicases, RNA helicases utilize energy derived from ATP hydrolysis to destabilize and unwind RNA duplexes. The most well-characterized and ubiquitous family of RNA helicases is the DEAD-box family, so named for the conserved B-type ATP-binding motif which is diagnostic of proteins in this family. Over 40 DEAD-box helicases have been identified in organisms as diverse as bacteria, insects, yeast, amphibians, mammals, and plants. DEAD-box helicases function in diverse processes  
10 such as translation initiation, splicing, ribosome assembly, and RNA editing, transport, and stability. Examples of these RNA helicases include yeast Drs1 protein, which is involved in ribosomal RNA processing; yeast TIF1 and TIF2 and mammalian eIF-4A, which are essential to the initiation of RNA translation; and human p68 antigen, which regulates cell growth and division (Ripmaster, T.L. et al. (1992) Proc. Natl. Acad. Sci. USA 89:11131-11135; Chang, T.-H. et al. (1990) Proc. Natl. Acad. Sci.  
15 USA 87:1571-1575). These RNA helicases demonstrate strong sequence homology over a stretch of some 420 amino acids. Included among these conserved sequences are the consensus sequence for the A motif of an ATP binding protein; the "DEAD box" sequence, associated with ATPase activity; the sequence SAT, associated with the actual helicase unwinding region; and an octapeptide consensus sequence, required for RNA binding and ATP hydrolysis (Pause, A. et al. (1993) Mol. Cell  
20 Biol. 13:6789-6798). Differences outside of these conserved regions are believed to reflect differences in the functional roles of individual proteins (Chang et al., *supra*).

Some DEAD-box helicases play tissue- and stage-specific roles in spermatogenesis and embryogenesis. Overexpression of the DEAD-box 1 protein (DDX1) may play a role in the progression of neuroblastoma (Nb) and retinoblastoma (Rb) tumors (Godbout, R. et al. (1998) J. Biol.  
25 Chem. 273:21161-21168). These observations suggest that DDX1 may promote or enhance tumor progression by altering the normal secondary structure and expression levels of RNA in cancer cells. Other DEAD-box helicases have been implicated either directly or indirectly in tumorigenesis (Godbout et al., *supra*). For example, murine p68 is mutated in ultraviolet light-induced tumors, and human DDX6 is located at a chromosomal breakpoint associated with B-cell lymphoma. Similarly, a  
30 chimeric protein comprised of DDX10 and NUP98, a nucleoporin protein, may be involved in the pathogenesis of certain myeloid malignancies.

### Topoisomerases

Besides the need to separate DNA strands prior to replication, the two strands must be

“unwound” from one another prior to their separation by DNA helicases. This function is performed by proteins known as DNA topoisomerases. DNA topoisomerase effectively acts as a reversible nuclease that hydrolyzes a phosphodiesterase bond in a DNA strand, permits the two strands to rotate freely about one another to remove the strain of the helix, and then rejoins the original phosphodiester bond between the two strands. Topoisomerases are essential enzymes responsible for the topological rearrangement of DNA brought about by transcription, replication, chromatin formation, recombination, and chromosome segregation. Superhelical coils are introduced into DNA by the passage of processive enzymes such as RNA polymerase, or by the separation of DNA strands by a helicase prior to replication. Knotting and concatenation can occur in the process of DNA synthesis, storage, and repair. All topoisomerases work by breaking a phosphodiester bond in the ribose-phosphate backbone of DNA. A catalytic tyrosine residue on the enzyme makes a nucleophilic attack on the scissile phosphodiester bond, resulting in a reaction intermediate in which a covalent bond is formed between the enzyme and one end of the broken strand. A tyrosine-DNA phosphodiesterase functions in DNA repair by hydrolyzing this bond in occasional dead-end topoisomerase I-DNA intermediates (Pouliot, J.J. et al. (1999) Science 286:552-555).

Two types of DNA topoisomerase exist, types I and II. Type I topoisomerases work as monomers, making a break in a single strand of DNA while type II topoisomerases, working as homodimers, cleave both strands. DNA Topoisomerase I causes a single-strand break in a DNA helix to allow the rotation of the two strands of the helix about the remaining phosphodiester bond in the opposite strand. DNA topoisomerase II causes a transient break in both strands of a DNA helix where two double helices cross over one another. This type of topoisomerase can efficiently separate two interlocked DNA circles (Alberts, *supra*, pp. 260-262). Type II topoisomerases are largely confined to proliferating cells in eukaryotes, such as cancer cells. For this reason they are targets for anticancer drugs. Topoisomerase II has been implicated in multi-drug resistance (MDR) as it appears to aid in the repair of DNA damage inflicted by DNA binding agents such as doxorubicin and vincristine.

The topoisomerase I family includes topoisomerases I and III (topo I and topo III). The crystal structure of human topoisomerase I suggests that rotation about the intact DNA strand is partially controlled by the enzyme. In this “controlled rotation” model, protein-DNA interactions limit the rotation, which is driven by torsional strain in the DNA (Stewart, L. et al. (1998) Science 379:1534-1541). Structurally, topo I can be recognized by its catalytic tyrosine residue and a number of other conserved residues in the active site region. Topo I is thought to function during transcription. Two topo IIIs are known in humans, and they are homologous to prokaryotic topoisomerase I, with a

conserved tyrosine and active site signature specific to this family. Topo III has been suggested to play a role in meiotic recombination. A mouse topo III is highly expressed in testis tissue and its expression increases with the increase in the number of cells in pachytene (Seki, T. et al. (1998) J. Biol. Chem. 273:28553-28556).

5           The topoisomerase II family includes two isozymes ( $\text{II}\alpha$  and  $\text{II}\beta$ ) encoded by different genes. Topo II cleaves double stranded DNA in a reproducible, nonrandom fashion, preferentially in an AT rich region, but the basis of cleavage site selectivity is not known. Structurally, topo II is made up of four domains, the first two of which are structurally similar and probably distantly homologous to similar domains in eukaryotic topo I. The second domain bears the catalytic tyrosine, as well as a  
10 highly conserved pentapeptide. The  $\text{II}\alpha$  isoform appears to be responsible for unlinking DNA during chromosome segregation. Cell lines expressing  $\text{II}\alpha$  but not  $\text{II}\beta$  suggest that  $\text{II}\beta$  is dispensable in cellular processes; however,  $\text{II}\beta$  knockout mice died perinatally due to a failure in neural development. That the major abnormalities occurred in predominantly late developmental events (neurogenesis) suggests that  $\text{II}\beta$  is needed not at mitosis, but rather during DNA repair (Yang, X. et al. (2000)  
15 Science 287:131-134).

Topoisomerases have been implicated in a number of disease states, and topoisomerase poisons have proven to be effective anti-tumor drugs for some human malignancies. Topo I is mislocalized in Fanconi's anemia, and may be involved in the chromosomal breakage seen in this disorder (Wunder, E. (1984) Hum. Genet. 68:276-281). Overexpression of a truncated topo III in  
20 ataxia-telangiectasia (A-T) cells partially suppresses the A-T phenotype, probably through a dominant negative mechanism. This suggests that topo III is deregulated in A-T (Fritz, E. et al. (1997) Proc. Natl. Acad. Sci. USA 94:4538-4542). Topo III also interacts with the Bloom's Syndrome gene product, and has been suggested to have a role as a tumor suppressor (Wu, L. et al. (2000) J. Biol. Chem. 275:9636-9644). Aberrant topo II activity is often associated with cancer or increased cancer  
25 risk. Greatly lowered topo II activity has been found in some, but not all A-T cell lines (Mohamed, R. et al. (1987) Biochem. Biophys. Res. Commun. 149:233-238). On the other hand, topo II can break DNA in the region of the A-T gene (ATM), which controls all DNA damage-responsive cell cycle checkpoints (Kaufmann, W.K. (1998) Proc. Soc. Exp. Biol. Med. 217:327-334). The ability of topoisomerases to break DNA has been used as the basis of antitumor drugs. Topoisomerase poisons  
30 act by increasing the number of dead-end covalent DNA-enzyme complexes in the cell, ultimately triggering cell death pathways (Fortune, J.M. and N. Osheroff (2000) Prog. Nucleic Acid Res. Mol. Biol. 64:221-253; Guichard, S.M. and M.K. Danks (1999) Curr. Opin. Oncol. 11:482-489). Antibodies against topo I are found in the serum of systemic sclerosis patients, and the levels of the antibody may

be used as a marker of pulmonary involvement in the disease (Diot, E. et al. (1999) *Chest* 116:715-720). Finally, the DNA binding region of human topo I has been used as a DNA delivery vehicle for gene therapy (Chen, T.Y. et al. (2000) *Appl. Microbiol. Biotechnol.* 53:558-567).

### Recombinases

5 Genetic recombination is the process of rearranging DNA sequences within an organism's genome to provide genetic variation for the organism in response to changes in the environment. DNA recombination allows variation in the particular combination of genes present in an individual's genome, as well as the timing and level of expression of these genes (Alberts, *supra*, pp. 263-273). Two broad classes of genetic recombination are commonly recognized, general recombination and  
10 site-specific recombination. General recombination involves genetic exchange between any homologous pair of DNA sequences usually located on two copies of the same chromosome. The process is aided by enzymes, recombinases, that "nick" one strand of a DNA duplex more or less randomly and permit exchange with a complementary strand on another duplex. The process does not normally change the arrangement of genes in a chromosome. In site-specific recombination, the  
15 recombinase recognizes specific nucleotide sequences present in one or both of the recombining molecules. Base-pairing is not involved in this form of recombination and therefore it does not require DNA homology between the recombining molecules. Unlike general recombination, this form of recombination can alter the relative positions of nucleotide sequences in chromosomes.

### RNA METABOLISM

20 Much of the regulation of gene expression in eucaryotic cells occurs at the posttranscriptional level. Messenger RNAs (mRNA), which are produced in the cell nucleus from primary transcripts of protein-encoding genes, are processed and transported to the cytoplasm where the protein synthesis machinery is located. RNA-binding proteins are a group of proteins that participate in the processing, editing, transport, localization, and posttranscriptional regulation of mRNAs, and comprise the protein  
25 component of ribosomes as well. The RNA-binding activity of many of these proteins is mediated by a series of RNA-binding motifs identified within them. These domains include the RNP motif, the arginine-rich motif, the RGG box, and the KH motif. (Reviewed in Burd, C.G. and Dreyfuss, G. (1994) *Science* 265:615-621.) The RNP motif is the most widely found and best characterized of these motifs. The RNP motif is composed of 90-100 amino acids which form an RNA-binding domain  
30 and is found in one or more copies in proteins that bind pre-mRNA, mRNA, pre-ribosomal RNA, and small nuclear RNAs. The RNP motif is composed of two short sequences (RNP-1 and RNP-2) and a number of other mostly hydrophobic, conserved amino acids interspersed throughout the motif. (Burd, *supra*; ExpASY PROSITE document PDOC0030.)

Ribonucleic acid (RNA) is a linear single-stranded polymer of four nucleotides, ATP, CTP, UTP, and GTP. In most organisms, RNA is transcribed as a copy of deoxyribonucleic acid (DNA), the genetic material of the organism. In retroviruses RNA rather than DNA serves as the genetic material. RNA copies of the genetic material encode proteins or serve various structural, catalytic, or regulatory roles in organisms. RNA is classified according to its cellular localization and function. Messenger RNAs (mRNAs) encode polypeptides. Ribosomal RNAs (rRNAs) are assembled, along with ribosomal proteins, into ribosomes, which are cytoplasmic particles that translate mRNA into polypeptides. Transfer RNAs (tRNAs) are cytosolic adaptor molecules that function in mRNA translation by recognizing both an mRNA codon and the amino acid that matches that codon. Heterogeneous nuclear RNAs (hnRNAs) include mRNA precursors and other nuclear RNAs of various sizes. Small nuclear RNAs (snRNAs) are a part of the nuclear spliceosome complex that removes intervening, non-coding sequences (introns) and rejoins exons in pre-mRNAs.

Proteins are associated with RNA during its transcription from DNA, RNA processing, and translation of mRNA into protein. Proteins are also associated with RNA as it is used for structural, catalytic, and regulatory purposes.

### RNA Processing

Ribosomal RNAs (rRNAs) are assembled, along with ribosomal proteins, into ribosomes, which are cytoplasmic particles that translate messenger RNA (mRNA) into polypeptides. The eukaryotic ribosome is composed of a 60S (large) subunit and a 40S (small) subunit, which together form the 80S ribosome. In addition to the 18S, 28S, 5S, and 5.8S rRNAs, ribosomes contain from 50 to over 80 different ribosomal proteins, depending on the organism. Ribosomal proteins are classified according to which subunit they belong (i.e., L, if associated with the large 60S large subunit or S if associated with the small 40S subunit). *E. coli* ribosomes have been the most thoroughly studied and contain 50 proteins, many of which are conserved in all life forms. The structures of nine ribosomal proteins have been solved to less than 3.0D resolution (i.e., S5, S6, S17, L1, L6, L9, L12, L14, L30), revealing common motifs, such as b-a-b protein folds in addition to acidic and basic RNA-binding motifs positioned between b-strands. Most ribosomal proteins are believed to contact rRNA directly (reviewed in Liljas, A. and M. Garber (1995) *Curr. Opin. Struct. Biol.* 5:721-727; see also Woodson, S.A. and N.B. Leontis (1998) *Curr. Opin. Struct. Biol.* 8:294-300; Ramakrishnan, V. and S.W. White (1998) *Trends Biochem. Sci.* 23:208-212).

Ribosomal proteins may undergo post-translational modifications or interact with other ribosome-associated proteins to regulate translation. For example, the highly homologous 40S ribosomal protein S6 kinases (S6K1 and S6K2) play a key role in the regulation of cell growth by

controlling the biosynthesis of translational components which make up the protein synthetic apparatus (including the ribosomal proteins). In the case of S6K1, at least eight phosphorylation sites are believed to mediate kinase activation in a hierarchical fashion (Dufner and Thomas (1999) *Exp. Cell. Res.* 253:100-109). Some of the ribosomal proteins, including L1, also function as translational  
5 repressors by binding to polycistronic mRNAs encoding ribosomal proteins (reviewed in Liljas and Garber, *supra*).

Recent evidence suggests that a number of ribosomal proteins have secondary functions independent of their involvement in protein biosynthesis. These proteins function as regulators of cell proliferation and, in some instances, as inducers of cell death. For example, the expression of human  
10 ribosomal protein L13a has been shown to induce apoptosis by arresting cell growth in the G2/M phase of the cell cycle. Inhibition of expression of L13a induces apoptosis in target cells, which suggests that this protein is necessary, in the appropriate amount, for cell survival. Similar results have been obtained in yeast where inactivation of yeast homologues of L13a, rp22 and rp23, results in severe growth retardation and death. A closely related ribosomal protein, L7, arrests cells in G1 and  
15 also induces apoptosis. Thus, it appears that a subset of ribosomal proteins may function as cell cycle checkpoints and compose a new family of cell proliferation regulators.

Mapping of individual ribosomal proteins on the surface of intact ribosomes is accomplished using 3D immunocryoelectronmicroscopy, whereby antibodies raised against specific ribosomal proteins are visualized. Progress has been made toward the mapping of L1, L7, and L12 while the  
20 structure of the intact ribosome has been solved to only 20-25D resolution and inconsistencies exist among different crude structures (Frank, J. (1997) *Curr. Opin. Struct. Biol.* 7:266-272).

Three distinct sites have been identified on the ribosome. The aminoacyl-tRNA acceptor site (A site) receives charged tRNAs (with the exception of the initiator-tRNA). The peptidyl-tRNA site (P site) binds the nascent polypeptide as the amino acid from the A site is added to the elongating  
25 chain. Deacylated tRNAs bind in the exit site (E site) prior to their release from the ribosome. (The structure of the ribosome is reviewed in Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY, pp. 888-908; Lodish, *supra*, pp. 119-138; and Lewin, B. (1997) Genes VI, Oxford University Press, Inc. New York NY).

Various proteins are necessary for processing of transcribed RNAs in the nucleus. Pre-  
30 mRNA processing steps include capping at the 5' end with methylguanosine, polyadenylating the 3' end, and splicing to remove introns. The primary RNA transcript from DNA is a faithful copy of the gene containing both exon and intron sequences, and the latter sequences must be cut out of the RNA transcript to produce a mRNA that codes for a protein. This "splicing" of the mRNA sequence takes

place in the nucleus with the aid of a large, multicomponent ribonucleoprotein complex known as a spliceosome. The spliceosomal complex is comprised of five small nuclear ribonucleoprotein particles (snRNPs) designated U1, U2, U4, U5, and U6. Each snRNP contains a single species of snRNA and about ten proteins. The RNA components of some snRNPs recognize and base-pair with intron  
5 consensus sequences. The protein components mediate spliceosome assembly and the splicing reaction. Autoantibodies to snRNP proteins are found in the blood of patients with systemic lupus erythematosus (Stryer, *supra*, p. 863).

Heterogeneous nuclear ribonucleoproteins (hnRNPs) have been identified that have roles in splicing, exporting of the mature RNAs to the cytoplasm, and mRNA translation (Biamonti, G. et al. (1998) Clin. Exp. Rheumatol. 16:317-326). Some examples of hnRNPs include the yeast proteins  
10 Hrp1p, involved in cleavage and polyadenylation at the 3' end of the RNA; Cbp80p, involved in capping the 5' end of the RNA; and Npl3p, a homolog of mammalian hnRNP A1, involved in export of mRNA from the nucleus (Shen, E.C. et al. (1998) Genes Dev. 12:679-691). HnRNPs have been shown to be important targets of the autoimmune response in rheumatic diseases (Biamonti et al.,  
15 *supra*).

Many snRNP and hnRNP proteins are characterized by an RNA recognition motif (RRM) (reviewed in Birney, E. et al. (1993) Nucleic Acids Res. 21:5803-5816). The RRM is about 80 amino acids in length and forms four  $\beta$ -strands and two  $\alpha$ -helices arranged in an  $\alpha/\beta$  sandwich. The RRM contains a core RNP-1 octapeptide motif along with surrounding conserved sequences. In addition to  
20 snRNP proteins, examples of RNA-binding proteins which contain the above motifs include heteronuclear ribonucleoproteins which stabilize nascent RNA and factors which regulate alternative splicing. Alternative splicing factors include developmentally regulated proteins, specific examples of which have been identified in lower eukaryotes such as *Drosophila melanogaster* and *Caenorhabditis elegans*. These proteins play key roles in developmental processes such as pattern  
25 formation and sex determination, respectively (Hodgkin, J. et al. (1994) Development 120:3681-3689).

The 3' ends of most eukaryote mRNAs are also posttranscriptionally modified by polyadenylation. Polyadenylation proceeds through two enzymatically distinct steps: (i) the endonucleolytic cleavage of nascent mRNAs at *cis*-acting polyadenylation signals in the 3'-untranslated (non-coding) region and (ii) the addition of a poly(A) tract to the 5' mRNA fragment.  
30 The presence of *cis*-acting RNA sequences is necessary for both steps. These sequences include 5'-AAUAAA-3' located 10-30 nucleotides upstream of the cleavage site and a less well-conserved GU- or U-rich sequence element located 10-30 nucleotides downstream of the cleavage site. Cleavage stimulation factor (CstF), cleavage factor I (CF I), and cleavage factor II (CF II) are involved in the

cleavage reaction while cleavage and polyadenylation specificity factor (CPSF) and poly(A) polymerase (PAP) are necessary for both cleavage and polyadenylation. An additional enzyme, poly(A)-binding protein II (PAB II), promotes poly(A) tract elongation (Rüegsegger, U. et al. (1996) *J. Biol. Chem.* 271:6107-6113; and references within).

## 5 TRANSLATION

Correct translation of the genetic code depends upon each amino acid forming a linkage with the appropriate transfer RNA (tRNA). The aminoacyl-tRNA synthetases (aaRSs) are essential proteins found in all living organisms. The aaRSs are responsible for the activation and correct attachment of an amino acid with its cognate tRNA, as the first step in protein biosynthesis.

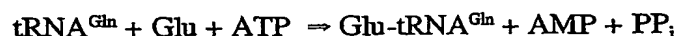
10 Prokaryotic organisms have at least twenty different types of aaRSs, one for each different amino acid, while eukaryotes usually have two aaRSs, a cytosolic form and a mitochondrial form, for each different amino acid. The 20 aaRS enzymes can be divided into two structural classes. Class I enzymes add amino acids to the 2' hydroxyl at the 3' end of tRNAs while Class II enzymes add amino acids to the 3' hydroxyl at the 3' end of tRNAs. Each class is characterized by a distinctive topology  
15 of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding Rossman 'fold'. In particular, a consensus tetrapeptide motif is highly conserved (Prosite Document PDOC00161, Aminoacyl-transfer RNA synthetases class-I signature). Class I enzymes are specific for arginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan, and valine. Class II enzymes contain a central catalytic domain, which consists of a seven-stranded  
20 antiparallel  $\beta$ -sheet domain, as well as N- and C- terminal regulatory domains. Class II enzymes are separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the latter group is further subdivided by the structure of the N- and C-terminal regulatory domains (Hartlein, M. and S. Cusack (1995) *J. Mol. Evol.* 40:519-530). Class II enzymes are specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine.

25 Certain aaRSs also have editing functions. IleRS, for example, can misactivate valine to form Val-tRNA<sup>Ile</sup>, but this product is cleared by a hydrolytic activity that destroys the mischarged product. This editing activity is located within a second catalytic site found in the connective polypeptide 1 region (CP1), a long insertion sequence within the Rossman fold domain of Class I enzymes (Schimmel, P. et al. (1998) *FASEB J.* 12:1599-1609). AaRSs also play a role in tRNA processing. It  
30 has been shown that mature tRNAs are charged with their respective amino acids in the nucleus before export to the cytoplasm, and charging may serve as a quality control mechanism to insure the tRNAs are functional (Martinis, S.A. et al. (1999) *EMBO J.* 18:4591-4596).

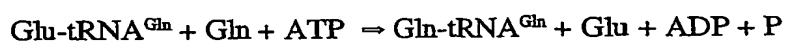
Under optimal conditions, polypeptide synthesis proceeds at a rate of approximately 40 amino

acid residues per second. The rate of misincorporation during translation is on the order of  $10^{-4}$  and is primarily the result of aminoacyl-t-RNAs being charged with the incorrect amino acid. Incorrectly charged tRNA are toxic to cells as they result in the incorporation of incorrect amino acid residues into an elongating polypeptide. The rate of translation is presumed to be a compromise between the optimal rate of elongation and the need for translational fidelity. Mathematical calculations predict that  $10^{-4}$  is indeed the maximum acceptable error rate for protein synthesis in a biological system (reviewed in Stryer, *supra*; and Watson, J. et al. (1987) The Benjamin/Cummings Publishing Co., Inc. Menlo Park, CA). A particularly error prone aminoacyl-tRNA charging event is the charging of tRNA<sup>Gln</sup> with Gln. A mechanism exists for the correction of this mischarging event which likely has its origins in evolution. Gln was among the last of the 20 naturally occurring amino acids used in polypeptide synthesis to appear in nature. Gram positive eubacteria, cyanobacteria, Archaeae, and eukaryotic organelles possess a noncanonical pathway for the synthesis of Gln-tRNA<sup>Gln</sup> based on the transformation of Glu-tRNA<sup>Gln</sup> (synthesized by Glu-tRNA synthetase, GluRS) using the enzyme Glu-tRNA<sup>Gln</sup> amidotransferase (Glu-AdT). The reactions involved in the transamidation pathway are as follows (Curnow, A.W. et al. (1997) Nucleic Acids Symposium 36:2-4):

#### GluRS



#### Glu-AdT



A similar enzyme, Asp-tRNA<sup>Asn</sup> amidotransferase, exists in Archaea, which transforms Asp-tRNA<sup>Asn</sup> to Asn-tRNA<sup>Asn</sup>. Formylase, the enzyme that transforms Met-tRNA<sup>fMet</sup> to fMet-tRNA<sup>fMet</sup> in eubacteria, is likely to be a related enzyme. A hydrolytic activity has also been identified that destroys mischarged Val-tRNA<sup>Ile</sup> (Schimmel, P. et al. (1998) FASEB J. 12:1599-1609). One likely scenario for the evolution of Glu-AdT in primitive life forms is the absence of a specific glutaminyl-tRNA synthetase (GlnRS), requiring an alternative pathway for the synthesis of Gln-tRNA<sup>Gln</sup>. In fact, deletion of the Glu-AdT operon in Gram positive bacteria is lethal (Curnow, A.W. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11819-11826). The existence of GluRS activity in other organisms has been inferred by the high degree of conservation in translation machinery in nature; however, GluRS has not been identified in all organisms, including *Homo sapiens*. Such an enzyme would be responsible for ensuring translational fidelity and reducing the synthesis of defective polypeptides.

In addition to their function in protein synthesis, specific aminoacyl tRNA synthetases also

play roles in cellular fidelity, RNA splicing, RNA trafficking, apoptosis, and transcriptional and translational regulation. For example, human tyrosyl-tRNA synthetase can be proteolytically cleaved into two fragments with distinct cytokine activities. The carboxy-terminal domain exhibits monocyte and leukocyte chemotaxis activity as well as stimulating production of myeloperoxidase, tumor  
 5 necrosis factor- $\alpha$ , and tissue factor. The N-terminal domain binds to the interleukin-8 type A receptor and functions as an interleukin-8-like cytokine. Human tyrosyl-tRNA synthetase is secreted from apoptotic tumor cells and may accelerate apoptosis (Wakasugi, K., and Schimmel, P. (1999) *Science* 284:147-151). Mitochondrial *Neurospora crassa* TyrRS and *S. cerevisiae* LeuRS are essential factors for certain group I intron splicing activities, and human mitochondrial LeuRS can substitute for  
 10 the yeast LeuRS in a yeast null strain. Certain bacterial aaRSs are involved in regulating their own transcription or translation (Martinis et al., *supra*). Several aaRSs are able to synthesize diadenosine oligophosphates, a class of signalling molecules with roles in cell proliferation, differentiation, and apoptosis (Kisselev, L.L et al. (1998) *FEBS Lett.* 427:157-163; Vartanian, A. et al. (1999) *FEBS Lett.* 456:175-180).

15 Autoantibodies against aminoacyl-tRNAs are generated by patients with autoimmune diseases such as rheumatic arthritis, dermatomyositis and polymyositis, and correlate strongly with complicating interstitial lung disease (ILD) (Freist, W. et al. (1999) *Biol. Chem.* 380:623-646; Freist, W. et al. (1996) *Biol. Chem. Hoppe Seyler* 377:343-356). These antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

20 Comparison of aaRS structures between humans and pathogens has been useful in the design of novel antibiotics (Schimmel et al., *supra*). Genetically engineered aaRSs have been utilized to allow site-specific incorporation of unnatural amino acids into proteins *in vivo* (Liu, D.R. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:10092-10097).

### tRNA Modifications

25 The modified ribonucleoside, pseudouridine ( $\psi$ ), is present ubiquitously in the anticodon regions of transfer RNAs (tRNAs), large and small ribosomal RNAs (rRNAs), and small nuclear RNAs (snRNAs).  $\psi$  is the most common of the modified nucleosides (i.e., other than G, A, U, and C) present in tRNAs. Only a few yeast tRNAs that are not involved in protein synthesis do not contain  $\psi$  (Cortese, R. et al. (1974) *J. Biol. Chem.* 249:1103-1108). The enzyme responsible for the conversion  
 30 of uridine to  $\psi$ , pseudouridine synthase (pseudouridylylase), was first isolated from *Salmonella typhimurium* (Arena, F. et al. (1978) *Nucleic Acids Res.* 5:4523-4536). The enzyme has since been isolated from a number of mammals, including steer and mice (Green, C.J. et al. (1982) *J. Biol. Chem.* 257:3045-52; and Chen, J. and J.R. Patton (1999) *RNA* 5:409-419). tRNA pseudouridine synthases

have been the most extensively studied members of the family. They require a thiol donor (e.g., cysteine) and a monovalent cation (e.g., ammonia or potassium) for optimal activity. Additional cofactors or high energy molecules (e.g., ATP or GTP) are not required (Green et al., *supra*). Other eukaryotic pseudouridine synthases have been identified that appear to be specific for rRNA (reviewed in Smith, C.M. and J.A. Steitz (1997) *Cell* 89:669-672) and a dual-specificity enzyme has been identified that uses both tRNA and rRNA substrates (Wrzesinski, J. et al. (1995) *RNA* 1: 437-448). The absence of  $\psi$  in the anticodon loop of tRNAs results in reduced growth in both bacteria (Singer, C.E. et al. (1972) *Nature New Biol.* 238:72-74) and yeast (Lecointe, F. (1998) *J. Biol. Chem.* 273:1316-1323), although the genetic defect is not lethal.

Another ribonucleoside modification that occurs primarily in eukaryotic cells is the conversion of guanosine to  $N^2,N^2$ -dimethylguanosine ( $m^2_2G$ ) at position 26 or 10 at the base of the D-stem of cytosolic and mitochondrial tRNAs. This posttranscriptional modification is believed to stabilize tRNA structure by preventing the formation of alternative tRNA secondary and tertiary structures. Yeast tRNA<sup>Asp</sup> is unusual in that it does not contain this modification. The modification does not occur in eubacteria, presumably because the structure of tRNAs in these cells and organelles is sequence constrained and does not require posttranscriptional modification to prevent the formation of alternative structures (Steinberg, S. and R. Cedergren (1995) *RNA* 1:886-891, and references within). The enzyme responsible for the conversion of guanosine to  $m^2_2G$  is a 63 kDa S-adenosylmethionine (SAM)-dependent tRNA  $N^2,N^2$ -dimethyl-guanosine methyltransferase (also referred to as the *TRM1* gene product and herein referred to as TRM) (Edqvist, J. (1995) *Biochimie* 77:54-61). The enzyme localizes to both the nucleus and the mitochondria (Li, J-M. et al. (1989) *J. Cell Biol.* 109:1411-1419). Based on studies with TRM from *Xenopus laevis*, there appears to be a requirement for base pairing at positions C11-G24 and G10-C25 immediately preceding the G26 to be modified, with other structural features of the tRNA also being required for the proper presentation of the G26 substrate (Edqvist, J. et al. (1992) *Nucleic Acids Res.* 20:6575-6581). Studies in yeast suggest that cells carrying a weak ochre tRNA suppressor (*sup3-i*) are unable to suppress translation termination in the absence of TRM activity, suggesting a role for TRM in modifying the frequency of suppression in eukaryotic cells (Niederberger, C. et al. (1999) *FEBS Lett.* 464:67-70), in addition to the more general function of ensuring the proper three-dimensional structures for tRNA.

### Translation Initiation

Initiation of translation can be divided into three stages. The first stage brings an initiator transfer RNA (Met-tRNA<sub>i</sub>) together with the 40S ribosomal subunit to form the 43S preinitiation complex. The second stage binds the 43S preinitiation complex to the mRNA, followed by migration

of the complex to the correct AUG initiation codon. The third stage brings the 60S ribosomal subunit to the 40S subunit to generate an 80S ribosome at the initiation codon. Regulation of translation primarily involves the first and second stage in the initiation process (Pain, V.M. (1996) Eur. J. Biochem. 236:747-771).

5        Several initiation factors, many of which contain multiple subunits, are involved in bringing an initiator tRNA and the 40S ribosomal subunit together. eIF2, a guanine nucleotide binding protein, recruits the initiator tRNA to the 40S ribosomal subunit. Only when eIF2 is bound to GTP does it associate with the initiator tRNA. eIF2B, a guanine nucleotide exchange protein, is responsible for converting eIF2 from the GDP-bound inactive form to the GTP-bound active form. Two other  
10       factors, eIF1A and eIF3 bind and stabilize the 40S subunit by interacting with the 18S ribosomal RNA and specific ribosomal structural proteins. eIF3 is also involved in association of the 40S ribosomal subunit with mRNA. The Met-tRNA<sub>i</sub>, eIF1A, eIF3, and 40S ribosomal subunit together make up the 43S preinitiation complex (Pain, *supra*).

      Additional factors are required for binding of the 43S preinitiation complex to an mRNA  
15       molecule, and the process is regulated at several levels. eIF4F is a complex consisting of three proteins: eIF4E, eIF4A, and eIF4G. eIF4E recognizes and binds to the mRNA 5'-terminal m<sup>7</sup>GTP cap, eIF4A is a bidirectional RNA-dependent helicase, and eIF4G is a scaffolding polypeptide. eIF4G has three binding domains. The N-terminal third of eIF4G interacts with eIF4E, the central third interacts with eIF4A, and the C-terminal third interacts with eIF3 bound to the 43S preinitiation  
20       complex. Thus, eIF4G acts as a bridge between the 40S ribosomal subunit and the mRNA (Hentze, M.W. (1997) Science 275:500-501).

      The ability of eIF4F to initiate binding of the 43S preinitiation complex is regulated by structural features of the mRNA. The mRNA molecule has an untranslated region (UTR) between the 5' cap and the AUG start codon. In some mRNAs this region forms secondary structures that  
25       impede binding of the 43S preinitiation complex. The helicase activity of eIF4A is thought to function in removing this secondary structure to facilitate binding of the 43S preinitiation complex (Pain, *supra*).

### **Translation Elongation**

      Elongation is the process whereby additional amino acids are joined to the initiator methionine  
30       to form the complete polypeptide chain. The elongation factors EF1 $\alpha$ , EF1 $\beta\gamma$ , and EF2 are involved in elongating the polypeptide chain following initiation. EF1 $\alpha$  is a GTP-binding protein. In EF1 $\alpha$ 's GTP-bound form, it brings an aminoacyl-tRNA to the ribosome's A site. The amino acid attached to the newly arrived aminoacyl-tRNA forms a peptide bond with the initiator methionine. The GTP on

EF1 $\alpha$  is hydrolyzed to GDP, and EF1 $\alpha$ -GDP dissociates from the ribosome. EF1 $\beta\gamma$  binds EF1 $\alpha$ -GDP and induces the dissociation of GDP from EF1 $\alpha$ , allowing EF1 $\alpha$  to bind GTP and a new cycle to begin.

As subsequent aminoacyl-tRNAs are brought to the ribosome, EF-G, another GTP-binding protein, catalyzes the translocation of tRNAs from the A site to the P site and finally to the E site of the ribosome. This allows the ribosome and the mRNA to remain attached during translation.

### Translation Termination

The release factor eRF carries out termination of translation. eRF recognizes stop codons in the mRNA, leading to the release of the polypeptide chain from the ribosome.

### Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

### Breast Cancer

There are more than 180,000 new cases of breast cancer diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (K. Gish (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752).

Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to

breast cancer and may be passed on from parents to children (Gish, K. (1999) AWIS Magazine 28:7-10). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast cancer is due to non-inherited mutations that occur in breast epithelial cells.

5           The relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied (see Khazaie, K. et al. (1993) Cancer and Metastasis Rev. 12:255-274, and references cited therein for a review of this area). Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor  
10 metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-  
15 2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S. S. et al. (1994) Am. J. Clin. Pathol. 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed in human breast carcinoma cells; Drg1 or RTP, a gene whose  
20 expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou, Z. et al. (1998) Int. J. Cancer 78:95-99; Chen, L. et al. (1990) Oncogene 5:1391-1395; Ulrix, W. et al (1999) FEBS Lett. 455:23-26; Sager, R. et al. (1996) Curr. Top. Microbiol. Immunol. 213:51-  
25 64; and Lee, S. W. et al. (1992) Proc. Natl. Acad. Sci. USA 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba, I.I. et al. (1998) Clin. Cancer Res. 4:2931-2938). Such a model is  
30 particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

The immune system responds to infection or trauma by activating a cascade of events that coordinate the progressive selection, amplification, and mobilization of cellular defense mechanisms.

A complex and balanced program of gene activation and repression is involved in this process. However, hyperactivity of the immune system as a result of improper or insufficient regulation of gene expression may result in considerable tissue or organ damage. This damage is well documented in immunological responses associated with arthritis, allergens, heart attack, stroke, and infections (Harrison's Principles of Internal Medicine, 13/e, McGraw Hill, Inc. and Teton Data Systems Software, 1996). In particular, a zinc finger protein termed Staf50 (for Stimulated trans-acting factor of 50 kDa) is a transcriptional regulator and is induced in various cell lines by interferon-I and -II. Staf50 appears to mediate the antiviral activity of interferon by down-regulating the viral transcription directed by the long terminal repeat promoter region of human immunodeficiency virus type-1 in transfected cells (Tissot, C. (1995) J. Biol. Chem. 270:14891-14898).

Dendritic cells (DC) are antigen presenting cells (APC) that play a key role in the primary immune response because of their unique ability to present antigens to naive T-cells. In addition, DC differentiate into separate subsets of mature immune cells that sustain and regulate immune responses following initial contact with antigen. DC subsets include those that preferentially induce particular T helper 1 (Th1) or T helper 2 (Th2) responses and those that regulate B cell responses. Moreover, DC are being used with increasing frequency to manipulate immune responses, either to downregulate aberrant autoimmune response or to enhance vaccination or tumor-specific response.

DC are functionally specialized in correlation with their particular differentiation state. CD34+ myeloid cells found in the bone marrow mature in response to signals into CD14+ CD11c+ monocytes. An innate or antigen non-specific response takes place initially when monocytes circulate to nonlymphoid tissues and respond to lipopolysaccharide (LPS), a bacterially-derived mitogen, and viruses. Such direct encounters with antigen cause secretion of pro-inflammatory cytokines that attract and regulate natural killer cells, macrophages, and eosinophils in the first line of defense against invading pathogens. Monocytes then mature into DC, which efficiently capture antigen through endocytosis and antigen-receptor uptake. Antigen processing and presentation trigger activation and differentiation into mature DC that express MHC class II molecules on the cell surface and efficiently activate T-cells, initiating antigen-specific T-cell and B-cell responses. In turn, T-cells activate DC through CD40 ligand - CD40 interactions, which stimulate expression of the costimulatory molecules CD80 and CD86, the latter most potent in amplifying T-cell responses. DC interaction via CD40 with T cells also stimulates the production of inflammatory cytokines such as TNF alpha and IL-1. Engagement of RANK, a member of the TNF receptor family by its ligand, TRANCE, which is expressed on activated T cells, enhances the survival of DC through inhibition of apoptosis, thereby enhancing T cell activation. The maturation and differentiation of monocytes into mature DC links the

antigen non-specific innate immune response to the antigen-specific adaptive immune response.

Human peripheral blood mononuclear cells (PBMCs) can be classified into discrete cellular populations representing the major components of the immune system. PBMCs contain about 52% lymphocytes (12% B lymphocytes, 40% T lymphocytes {25% CD4+ and 15% CD8+}), 20% NK  
5 cells, 25% monocytes, and 3% various cells that include dendritic cells and progenitor cells. The proportions, as well as the biology of these cellular components tend to vary slightly between healthy individuals, depending on factors such as age, past medical history, and genetic backgrounds.

### Steroid Hormones

Steroids are a class of lipid-soluble molecules, including cholesterol, bile acids, vitamin D, and  
10 hormones, that share a common four-ring structure based on cyclopentanoperhydrophenanthrene and that carry out a wide variety of functions. Cholesterol, for example, is a component of cell membranes that controls membrane fluidity. It is also a precursor for bile acids which solubilize lipids and facilitate absorption in the small intestine during digestion. Vitamin D regulates the absorption of calcium in the small intestine and controls the concentration of calcium in plasma. Steroid hormones,  
15 produced by the adrenal cortex, ovaries, and testes, include glucocorticoids, mineralocorticoids, androgens, and estrogens. They control various biological processes by binding to intracellular receptors that regulate transcription of specific genes in the nucleus. Glucocorticoids, for example, increase blood glucose concentrations by regulation of gluconeogenesis in the liver, increase blood concentrations of fatty acids by promoting lipolysis in adipose tissues, modulate sensitivity to  
20 catecholamines in the central nervous system, and reduce inflammation. The principal mineralocorticoid, aldosterone, is produced by the adrenal cortex and acts on cells of the distal tubules of the kidney to enhance sodium ion reabsorption. Androgens, produced by the interstitial cells of Leydig in the testis, include the male sex hormone testosterone, which triggers changes at puberty, the production of sperm and maintenance of secondary sexual characteristics. Female sex hormones,  
25 estrogen and progesterone, are produced by the ovaries and also by the placenta and adrenal cortex of the fetus during pregnancy. Estrogen regulates female reproductive processes and secondary sexual characteristics. Progesterone regulates changes in the endometrium during the menstrual cycle and pregnancy.

Steroid hormones are widely used for fertility control and in anti-inflammatory treatments for  
30 physical injuries and diseases such as arthritis, asthma, and auto-immune disorders. Progesterone, a naturally occurring progestin, is primarily used to treat amenorrhea, abnormal uterine bleeding, or as a contraceptive. Endogenous progesterone is responsible for inducing secretory activity in the endometrium of the estrogen-primed uterus in preparation for the implantation of a fertilized egg and

for the maintenance of pregnancy. It is secreted from the corpus luteum in response to luteinizing hormone (LH). The primary contraceptive effect of exogenous progestins involves the suppression of the midcycle surge of LH. At the cellular level, progestins diffuse freely into target cells and bind to the progesterone receptor. Target cells include the female reproductive tract, the mammary gland, 5 the hypothalamus, and the pituitary. Once bound to the receptor, progestins slow the frequency of release of gonadotropin releasing hormone from the hypothalamus and blunt the pre-ovulatory LH surge, thereby preventing follicular maturation and ovulation. Progesterone has minimal estrogenic and androgenic activity. Progesterone is metabolized hepatically to pregnanediol and conjugated with glucuronic acid.

10 Medroxyprogesterone (MAH), also known as 6 $\alpha$ -methyl-17-hydroxyprogesterone, is a synthetic progestin with a pharmacological activity about 15 times greater than progesterone. MAH is used for the treatment of renal and endometrial carcinomas, amenorrhea, abnormal uterine bleeding, and endometriosis associated with hormonal imbalance. MAH has a stimulatory effect on respiratory centers and has been used in cases of low blood oxygenation caused by sleep apnea, chronic 15 obstructive pulmonary disease, or hypercapnia.

Mifepristone, also known as RU-486, is an antiprogesterone drug that blocks receptors of progesterone. It counteracts the effects of progesterone, which is needed to sustain pregnancy. Mifepristone induces spontaneous abortion when administered in early pregnancy followed by treatment with the prostaglandin, misoprostol. Further, studies show that mifepristone at a 20 substantially lower dose can be highly effective as a postcoital contraceptive when administered within five days after unprotected intercourse, thus providing women with a "morning-after pill" in case of contraceptive failure or sexual assault. Mifepristone also has potential uses in the treatment of breast and ovarian cancers in cases in which tumors are progesterone-dependent. It interferes with steroid-dependent growth of brain meningiomas, and may be useful in treatment of endometriosis where it 25 blocks the estrogen-dependent growth of endometrial tissues. It may also be useful in treatment of uterine fibroid tumors and Cushing's Syndrome. Mifepristone binds to glucocorticoid receptors and interferes with cortisol binding. Mifepristone also may act as an anti-glucocorticoid and be effective for treating conditions where cortisol levels are elevated such as AIDS, anorexia nervosa, ulcers, diabetes, Parkinson's disease, multiple sclerosis, and Alzheimer's disease.

30 Danazol is a synthetic steroid derived from ethinyl testosterone. Danazol indirectly reduces estrogen production by lowering pituitary synthesis of follicle-stimulating hormone and LH. Danazol also binds to sex hormone receptors in target tissues, thereby exhibiting anabolic, antiestrogenic, and weakly androgenic activity. Danazol does not possess any progestogenic activity, and does not

suppress normal pituitary release of corticotropin or release of cortisol by the adrenal glands. Danazol is used in the treatment of endometriosis to relieve pain and inhibit endometrial cell growth. It is also used to treat fibrocystic breast disease and hereditary angioedema.

Corticosteroids are used to relieve inflammation and to suppress the immune response. They inhibit eosinophil, basophil, and airway epithelial cell function by regulation of cytokines that mediate the inflammatory response. They inhibit leukocyte infiltration at the site of inflammation, interfere in the function of mediators of the inflammatory response, and suppress the humoral immune response. Corticosteroids are used to treat allergies, asthma, arthritis, and skin conditions. Beclomethasone is a synthetic glucocorticoid that is used to treat steroid-dependent asthma, to relieve symptoms associated with allergic or nonallergic (vasomotor) rhinitis, or to prevent recurrent nasal polyps following surgical removal. The anti-inflammatory and vasoconstrictive effects of intranasal beclomethasone are 5000 times greater than those produced by hydrocortisone. Budesonide is a corticosteroid used to control symptoms associated with allergic rhinitis or asthma. Budesonide has high topical anti-inflammatory activity but low systemic activity. Dexamethasone is a synthetic glucocorticoid used in anti-inflammatory or immunosuppressive compositions. It is also used in inhalants to prevent symptoms of asthma. Due to its greater ability to reach the central nervous system, dexamethasone is usually the treatment of choice to control cerebral edema. Dexamethasone is approximately 20-30 times more potent than hydrocortisone and 5-7 times more potent than prednisone. Prednisone is metabolized in the liver to its active form, prednisolone, a glucocorticoid with anti-inflammatory properties. Prednisone is approximately 4 times more potent than hydrocortisone and the duration of action of prednisone is intermediate between hydrocortisone and dexamethasone. Prednisone is used to treat allograft rejection, asthma, systemic lupus erythematosus, arthritis, ulcerative colitis, and other inflammatory conditions. Betamethasone is a synthetic glucocorticoid with antiinflammatory and immunosuppressive activity and is used to treat psoriasis and fungal infections, such as athlete's foot and ringworm.

The anti-inflammatory actions of corticosteroids are thought to involve phospholipase A<sub>2</sub> inhibitory proteins, collectively called lipocortins. Lipocortins, in turn, control the biosynthesis of potent mediators of inflammation such as prostaglandins and leukotrienes by inhibiting the release of the precursor molecule arachidonic acid. Proposed mechanisms of action include decreased IgE synthesis, increased number of  $\beta$ -adrenergic receptors on leukocytes, and decreased arachidonic acid metabolism. During an immediate allergic reaction, such as in chronic bronchial asthma, allergens bridge the IgE antibodies on the surface of mast cells, which triggers these cells to release chemotactic substances. Mast cell influx and activation, therefore, is partially responsible for the

inflammation and hyperirritability of the oral mucosa in asthmatic patients. This inflammation can be retarded by administration of corticosteroids.

There is a need in the art for new compositions, including nucleic acids and proteins, for the  
5 diagnosis, prevention, and treatment of cell proliferative, neurological, reproductive, developmental, autoimmune/inflammatory, and DNA repair disorders, and infections.

## SUMMARY OF THE INVENTION

10 Various embodiments of the invention provide purified polypeptides, nucleic acid-associated proteins, referred to collectively as "NAAP" and individually as "NAAP-1," "NAAP-2," "NAAP-3," "NAAP-4," "NAAP-5," "NAAP-6," "NAAP-7," "NAAP-8," "NAAP-9," "NAAP-10," "NAAP-11," "NAAP-12," "NAAP-13," "NAAP-14," "NAAP-15," "NAAP-16," "NAAP-17," "NAAP-18," "NAAP-19," "NAAP-20," "NAAP-21," "NAAP-22," "NAAP-23," "NAAP-24," "NAAP-25,"  
15 "NAAP-26," "NAAP-27," "NAAP-28," "NAAP-29," "NAAP-30," "NAAP-31," "NAAP-32," "NAAP-33," "NAAP-34," "NAAP-35," and "NAAP-36," and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified nucleic acid-associated proteins and/or their encoding polynucleotides for facilitating the drug discovery process, including  
20 determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified nucleic acid-associated proteins and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at  
25 least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36. Another  
30 embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-36.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, b) a polypeptide comprising a naturally occurring amino acid

sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-36. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:37-72.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, c) a biologically active fragment of

a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-72, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-72, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-72, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-72, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-72, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-72, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain

reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide  
5 selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36,  
10 and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-36. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional NAAP, comprising administering to a  
15 patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an  
20 amino acid sequence selected from the group consisting of SEQ ID NO:1-36, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another  
25 embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional NAAP, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness  
30 as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, c) a

biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional NAAP, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-72, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-72, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-72, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-72, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-72, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank

homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide embodiments, along with allele frequencies in different human populations.

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an

admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

“NAAP” refers to the amino acid sequences of substantially purified NAAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term “agonist” refers to a molecule which intensifies or mimics the biological activity of NAAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NAAP either by directly interacting with NAAP or by acting on components of the biological pathway in which NAAP participates.

An “allelic variant” is an alternative form of the gene encoding NAAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding NAAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as NAAP or a polypeptide with at least one functional characteristic of NAAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding NAAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding NAAP. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent NAAP. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of NAAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms “amino acid” and “amino acid sequence” can refer to an oligopeptide, a peptide, a

polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where “amino acid sequence” is recited to refer to a sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

5 “Amplification” relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of NAAP. Antagonists may include proteins such as antibodies, anticalins, nucleic acids,  
10 carbohydrates, small molecules, or any other compound or composition which modulates the activity of NAAP either by directly interacting with NAAP or by acting on components of the biological pathway in which NAAP participates.

The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant.  
15 Antibodies that bind NAAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and  
20 keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term “antigenic determinant” refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on  
25 the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term “aptamer” refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No.  
30 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a

ribonucleotide may be replaced by 2'-F or 2'-NH<sub>2</sub>), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic NAAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given

polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding NAAP or fragments of NAAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

5           A “deletion” refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term “derivative” refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains  
10           at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A “detectable label” refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

15           “Differential expression” refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

20           “Exon shuffling” refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A “fragment” is a unique portion of NAAP or a polynucleotide encoding NAAP which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up  
25           to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule.  
30           For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present

embodiments.

A fragment of SEQ ID NO:37-72 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:37-72, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:37-72 can be employed  
5 in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:37-72 from related polynucleotides. The precise length of a fragment of SEQ ID NO:37-72 and the region of SEQ ID NO:37-72 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

10 A fragment of SEQ ID NO:1-36 is encoded by a fragment of SEQ ID NO:37-72. A fragment of SEQ ID NO:1-36 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-36. For example, a fragment of SEQ ID NO:1-36 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-36. The precise length of a fragment of SEQ ID NO:1-36 and the region of SEQ ID NO:1-36 to which  
15 the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A “full length” polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A “full length” polynucleotide sequence encodes a “full length” polypeptide sequence.

20 “Homology” refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in  
25 the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into  
30 the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins and Sharp (1989; CABIOS 5:151-153) and in Higgins et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the

default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes

in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases “percent identity” and “% identity,” as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and “diagonals saved”=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain

DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook et al. (1989; Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily

apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of NAAP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of NAAP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of NAAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of NAAP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably

linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

“Peptide nucleic acid” (PNA) refers to an antisense molecule or anti-gene agent which  
5 comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

“Post-translational modification” of an NAAP may involve lipidation, glycosylation,  
10 phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of NAAP.

“Probe” refers to nucleic acids encoding NAAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides  
15 or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. “Primers” are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic  
20 acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100,  
25 or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al. (*supra*); Ausubel, F.M. et al. (1987; Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY); and Innis, M. et al. (1990; PCR  
30 Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is

expressed, inducing a protective immunological response in the mammal.

A “regulatory element” refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription,  
5 translation, or RNA stability.

“Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

10 An “RNA equivalent,” in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term “sample” is used in its broadest sense. A sample suspected of containing NAAP,  
15 nucleic acids encoding NAAP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms “specific binding” and “specifically binding” refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or  
20 synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope “A,” the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

25 The term “substantially purified” refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A “substitution” refers to the replacement of one or more amino acid residues or nucleotides  
30 by different amino acid residues or nucleotides, respectively.

“Substrate” refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells,

trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A “transcript image” or “expression profile” refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

“Transformation” describes a process by which exogenous DNA is introduced into a recipient  
5 cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral  
10 infection, electroporation, heat shock, lipofection, and particle bombardment. The term “transformed cells” includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A “transgenic organism,” as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid  
15 introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The  
20 term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques  
25 for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al., *supra*.

A “variant” of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-  
30 1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an

“allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass “single nucleotide polymorphisms” (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A “variant” of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

Various embodiments of the invention include new human nucleic acid-associated proteins (NAAP), the polynucleotides encoding NAAP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, neurological, reproductive, developmental, autoimmune/inflammatory, and DNA repair disorders, and infections.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to polypeptide and polynucleotide embodiments. The full length clones encode polypeptides which have at least 95%

sequence identity to the polypeptides shown in column 3.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are nucleic acid-associated proteins. For example, SEQ ID NO:1 is 88% identical, from residue M1 to residue L304, to mouse genomic screen homeobox protein 2 (GenBank ID g1042009) as determined by the Basic Local Alignment Search Tool (BLAST). The BLAST probability score is  $2.1e-146$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a homeobox domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, PROFILESCAN, and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:1 is a homeobox protein.

In an alternative example, SEQ ID NO:8 is 99% identical, from residue G24 to residue E384, to human DNA-binding protein B (GenBank ID g181486) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $5.8e-199$ , which

indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 also contains 'cold-shock' DNA-binding domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:8 is a DNA-binding protein.

In another example, SEQ ID NO:13 is 94% identical, from residue M780 to residue E1598, to human centriole associated protein CEP110 (GenBank ID g3435244) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from MOTIFS and further BLAST analyses provide corroborative evidence that SEQ ID NO:13 is a centriole associated protein.

In yet another example, SEQ ID NO:15 is 87% identical, from residue E435 to residue L2523, to human bromodomain PHD finger transcription factor (GenBank ID g6683492) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:15 also contains a PHD finger domain and a bromodomain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:15 is a bromodomain PHD finger transcription factor.

In an alternative example, SEQ ID NO:21 is 41% identical, from residue T141 to residue E370, to human Kurppel-like zinc finger protein HZF2 (GenBank ID g8163824) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $6.7e-71$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:21 also contains Zinc finger C2H2 type domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:21 is a C2H2 type zinc finger protein.

In yet another example, SEQ ID NO:30 is 33% identical, from residue T556 to residue E1699, and 32% identical, from residue S10 to Y211, to *Schizosaccharomyces pombe* putative helicase (GenBank ID g6901197) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $2.9e-137$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:30 also contains a DEAD/DEAH

box helicase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST analyses of the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:30 is a helicase.

5 In yet another example, SEQ ID NO:33 is 97% identical, from residue M1 to residue V602, a murine T-box transcription factor (GenBank ID g3169261) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:33 also contains a T-box domain as determined by searching for statistically significant  
10 matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, PRODOM and DOMO BLAST, and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:33 is a transcription factor molecule.

Taken together, the foregoing provides evidence that SEQ ID NO:1, SEQ ID NO:8, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:21, SEQ ID NO:30, and SEQ ID NO:33 are all molecules  
15 associated with nucleic acids. SEQ ID NO:2-7, SEQ ID NO:9-12, SEQ ID NO:14, SEQ ID NO:16-20, SEQ ID NO:22-29, SEQ ID NO:31-32, and SEQ ID NO:34-36 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-36 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA  
20 sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic  
25 sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:37-72 or that distinguish between SEQ ID NO:37-72 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA  
30 libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation

“ENST”). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation “NM” or “NT”) or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation “NP”). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon stitching” algorithm. For example, a polynucleotide sequence identified as FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a “stitched” sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3...</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an “exon-stretching” algorithm. For example, a polynucleotide sequence identified as FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is a “stretched” sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the “exon-stretching” algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the “exon-stretching” algorithm, a RefSeq identifier (denoted by “NM,” “NP,” or “NT”) may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide embodiments, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses NAAP variants. A preferred NAAP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the NAAP amino acid sequence, and which contains at least one functional or structural characteristic of NAAP.

Various embodiments also encompass polynucleotides which encode NAAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:37-72, which encodes NAAP. The polynucleotide sequences of SEQ ID NO:37-72, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding NAAP. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even

at least about 95% polynucleotide sequence identity to a polynucleotide encoding NAAP. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:37-72 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:37-72. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of NAAP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding NAAP. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding NAAP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding NAAP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding NAAP. For example, a polynucleotide comprising a sequence of SEQ ID NO:72 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:50. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of NAAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding NAAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring NAAP, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode NAAP and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring NAAP under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding NAAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a

particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding NAAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode NAAP and NAAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding NAAP or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:37-72 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding NAAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a

cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture  
5 PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al.  
10 (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be  
15 about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T)  
20 library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the  
25 emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments  
30 which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode NAAP may be cloned in recombinant DNA molecules that direct expression of NAAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the

genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express NAAP.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter NAAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of NAAP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding NAAP may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, NAAP itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of NAAP, or any part thereof, may be altered during direct

synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active NAAP, the polynucleotides encoding NAAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding NAAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding NAAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding NAAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding NAAP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook et al., *supra*, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding NAAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors

(e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook et al., *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding NAAP. For example, routine cloning, subcloning, and propagation of polynucleotides encoding NAAP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding NAAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of NAAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of NAAP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of NAAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel, 1995, *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of NAAP. Transcription of polynucleotides encoding NAAP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used

alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These  
5 constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding NAAP may be ligated  
10 into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses NAAP in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors  
15 may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of  
20 NAAP in cell lines is preferred. For example, polynucleotides encoding NAAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched  
25 media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These  
30 include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *ap<sup>r</sup>* cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to

methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding NAAP is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding NAAP can be identified by the absence of marker gene function.

Alternatively, a marker gene can be placed in tandem with a sequence encoding NAAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding NAAP and that express NAAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of NAAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on NAAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and

may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding NAAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide.

Alternatively, polynucleotides encoding NAAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding NAAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode NAAP may be designed to contain signal sequences which direct secretion of NAAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding NAAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric NAAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of NAAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose

binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the NAAP encoding sequence and the heterologous protein sequence, so that NAAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled NAAP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

NAAP, fragments of NAAP, or variants of NAAP may be used to screen for compounds that specifically bind to NAAP. One or more test compounds may be screened for specific binding to NAAP. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to NAAP. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of NAAP can be used to screen for binding of test compounds, such as antibodies, to NAAP, a variant of NAAP, or a combination of NAAP and/or one or more variants NAAP. In an embodiment, a variant of NAAP can be used to screen for compounds that bind to a variant of NAAP, but not to NAAP having the exact sequence of a sequence of SEQ ID NO:1-36. NAAP variants used to perform such screening can have a range of about 50% to about 99% sequence identity to NAAP, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to NAAP can be closely related to the natural ligand of NAAP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor NAAP (Howard, A.D. et al. (2001) *Trends Pharmacol. Sci.* 22:132-140; Wise, A. et al. (2002) *Drug Discovery Today* 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to NAAP can be closely related to the natural receptor to which NAAP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for NAAP which is capable of propagating a signal, or a  
5 decoy receptor for NAAP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Immunex Corp., Seattle WA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75  
10 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG<sub>1</sub> (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to NAAP, fragments of NAAP, or variants of NAAP. The binding specificity of the antibodies thus screened can thereby be selected to identify  
15 particular fragments or variants of NAAP. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of NAAP. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of NAAP.

In an embodiment, anticalins can be screened for specific binding to NAAP, fragments of NAAP, or variants of NAAP. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel  
20 having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or  
25 significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit  
30 NAAP involves producing appropriate cells which express NAAP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing NAAP or cell membrane fractions which contain NAAP are then contacted with a test

compound and binding, stimulation, or inhibition of activity of either NAAP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with NAAP, either in solution or affixed to a solid support, and detecting the binding of NAAP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

NAAP, fragments of NAAP, or variants of NAAP may be used to screen for compounds that modulate the activity of NAAP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for NAAP activity, wherein NAAP is combined with at least one test compound, and the activity of NAAP in the presence of a test compound is compared with the activity of NAAP in the absence of the test compound. A change in the activity of NAAP in the presence of the test compound is indicative of a compound that modulates the activity of NAAP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising NAAP under conditions suitable for NAAP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of NAAP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding NAAP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES

cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding NAAP may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding NAAP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding NAAP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress NAAP, e.g., by secreting NAAP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of NAAP and nucleic acid-associated proteins. In addition, examples of tissues expressing NAAP can be found in Table 6 and can also be found in Example XI. Therefore, NAAP appears to play a role in cell proliferative, neurological, reproductive, developmental, autoimmune/inflammatory, and DNA repair disorders, and infections. In the treatment of disorders associated with increased NAAP expression or activity, it is desirable to decrease the expression or activity of NAAP. In the treatment of disorders associated with decreased NAAP expression or activity, it is desirable to increase the expression or activity of NAAP.

Therefore, in one embodiment, NAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed  
5 connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin,  
10 spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as progressive supranuclear palsy, corticobasal degeneration, familial frontotemporal dementia, epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary  
15 ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal  
20 hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorder of the central nervous system, cerebral palsy, a neuroskeletal disorder, an autonomic nervous system disorder, a cranial nerve disorder, a spinal cord disease, muscular dystrophy and other neuromuscular disorder, a peripheral nervous system disorder, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathy, myasthenia gravis, periodic paralysis, a mental  
25 disorder including mood, anxiety, and schizophrenic disorder, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation  
30 syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast,

and gynecomastia; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; an infection, such as those caused by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, or togavirus; an infection caused by a bacterial agent classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, or campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, or mycoplasma; an infection caused by a fungal agent classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malassezia, histoplasma, or other mycosis-causing fungal agent; an infection caused by a parasite classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematode such as trichinella, intestinal nematode such as ascaris, lymphatic filarial nematode, trematode such as

schistosoma, and cestode such as tapeworm; and a DNA repair disorder such as xeroderma pigmentosum, Bloom's syndrome, and Werner's syndrome.

In another embodiment, a vector capable of expressing NAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified NAAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of NAAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of NAAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NAAP. Examples of such disorders include, but are not limited to, those cell proliferative, neurological, reproductive, developmental, autoimmune/inflammatory, and DNA repair disorders, and infections, described above. In one aspect, an antibody which specifically binds NAAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express NAAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding NAAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NAAP including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of NAAP may be produced using methods which are generally known in the art. In particular, purified NAAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind NAAP. Antibodies to NAAP may also be generated using methods that are well known in the art. Such antibodies may include, but are

not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with NAAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to NAAP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of NAAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to NAAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce NAAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated

by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for NAAP may also be generated. For example, such fragments include, but are not limited to,  $F(ab')_2$  fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the  $F(ab')_2$  fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between NAAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering NAAP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for NAAP. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of NAAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple NAAP epitopes, represents the average affinity, or avidity, of the antibodies for NAAP. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular NAAP epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the NAAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of NAAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons,

New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of NAAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*, and Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding NAAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding NAAP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding NAAP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) *Blood* 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736).

In another embodiment of the invention, polynucleotides encoding NAAP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene*

Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in NAAP expression or regulation causes disease, the expression of NAAP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in NAAP are treated by constructing mammalian expression vectors encoding NAAP and introducing these vectors by mechanical means into NAAP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of NAAP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). NAAP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding NAAP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to NAAP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding NAAP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding NAAP to cells which have one or more genetic abnormalities with respect to the expression of NAAP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are

described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma and Somia (1997; Nature 18:389:239-242), both incorporated by reference herein.

5 In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding NAAP to target cells which have one or more genetic abnormalities with respect to the expression of NAAP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing NAAP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with  
10 ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92  
15 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, et al. (1999; J. Virol. 73:519-532) and Xu et al. (1994; Dev. Biol. 163:152-161), hereby incorporated by reference. The manipulation of cloned herpesvirus  
20 sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to  
25 deliver polynucleotides encoding NAAP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA,  
30 resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for NAAP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of NAAP-coding RNAs and the synthesis of high levels of NAAP in vector transduced cells. While

alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of NAAP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding NAAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively,

RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding NAAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

5 RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine,  
10 and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding NAAP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not  
15 limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased NAAP  
20 expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding NAAP may be therapeutically useful, and in the treatment of disorders associated with decreased NAAP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding NAAP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in  
25 altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a  
30 library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding NAAP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding NAAP are assayed by

any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding NAAP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of NAAP, antibodies to NAAP, and mimetics, agonists, antagonists, or inhibitors of NAAP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical,

sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-  
5 acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

10 Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising NAAP or fragments thereof. For example, liposome preparations  
15 containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, NAAP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

20 For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

25 A therapeutically effective dose refers to that amount of active ingredient, for example NAAP or fragments thereof, antibodies of NAAP, and agonists, antagonists or inhibitors of NAAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) or LD<sub>50</sub> (the dose  
30 lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD<sub>50</sub>/ED<sub>50</sub> ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is

preferably within a range of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu$ g to 100,000  $\mu$ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind NAAP may be used for the diagnosis of disorders characterized by expression of NAAP, or in assays to monitor patients being treated with NAAP or agonists, antagonists, or inhibitors of NAAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for NAAP include methods which utilize the antibody and a label to detect NAAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring NAAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of NAAP expression. Normal or standard values for NAAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to NAAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of NAAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding NAAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of NAAP may be correlated with disease. The  
5 diagnostic assay may be used to determine absence, presence, and excess expression of NAAP, and to monitor regulation of NAAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding NAAP or closely related molecules may be used to identify nucleic acid sequences which encode NAAP. The specificity of the probe, whether it is made from a  
10 highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding NAAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the NAAP encoding sequences. The hybridization probes of the subject  
15 invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:37-72 or from genomic sequences including promoters, enhancers, and introns of the NAAP gene.

Means for producing specific hybridization probes for polynucleotides encoding NAAP include the cloning of polynucleotides encoding NAAP or NAAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to  
20 synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding NAAP may be used for the diagnosis of disorders associated with  
25 expression of NAAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal  
30 gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as progressive supranuclear palsy, corticobasal degeneration, familial frontotemporal dementia, epilepsy, ischemic cerebrovascular

disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorder of the central nervous system, cerebral palsy, a neuroskeletal disorder, an autonomic nervous system disorder, a cranial nerve disorder, a spinal cord disease, muscular dystrophy and other neuromuscular disorder, a peripheral nervous system disorder, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathy, myasthenia gravis, periodic paralysis, a mental disorder including mood, anxiety, and schizophrenic disorder, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis,

cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis,

5 myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral,

10 bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; an infection, such as those caused by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, or togavirus; an infection caused by a bacterial agent classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus,

15 legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, or campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, or mycoplasma; an infection caused by a fungal agent classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malassezia, histoplasma, or other mycosis-causing fungal agent; an infection caused by a parasite classified as

20 plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematode such as trichinella, intestinal nematode such as ascaris, lymphatic filarial nematode, trematode such as schistosoma, and cestode such as tapeworm; and a DNA repair disorder such as xeroderma pigmentosum, Bloom's syndrome, and Werner's syndrome. Polynucleotides encoding NAAP may be

25 used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered NAAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, polynucleotides encoding NAAP may be used in assays that detect the

30 presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding NAAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a

standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding NAAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of NAAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding NAAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding NAAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding NAAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding NAAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding NAAP

may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding NAAP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of NAAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from

standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid  
5 quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic  
10 variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment  
15 regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, NAAP, fragments of NAAP, or antibodies specific for NAAP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to  
20 generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No.  
25 5,840,484, expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The  
30 resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed  
5 molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and  
10 gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in  
15 interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

20 In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in  
25 an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually  
30 to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel

electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for NAAP to quantify the levels of NAAP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendozze, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the

test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London.

In another embodiment of the invention, nucleic acid sequences encoding NAAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM)

World Wide Web site. Correlation between the location of the gene encoding NAAP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

*In situ* hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, NAAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between NAAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with NAAP, or fragments thereof, and washed. Bound NAAP is then detected by methods well known in the art. Purified NAAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding NAAP specifically compete with a test compound for binding NAAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with NAAP.

In additional embodiments, the nucleotide sequences which encode NAAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such

properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/301,893, U.S. Ser. No. 60/300,518, U.S. Ser. No. 60/301,787, U.S. Ser. No. 60/301,892, U.S. Ser. No. 60/301,792, U.S. Ser. No. 60/303,442, U.S. Ser. No. 60/303,405, and U.S. Ser. No. 60/364,438, are expressly incorporated by reference herein.

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel, 1997, *supra*, units 5.1-6.6). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography

(Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPO1 plasmid (Invitrogen), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Invitrogen.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4 °C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows.

Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard

ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

5           The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and  
10       BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) *Nucleic Acids Res.* 29:41-43); and HMM-based protein domain databases such as  
15       SMART (Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:5857-5864; Letunic, I. et al. (2002) *Nucleic Acids Res.* 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365). The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide  
20       sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated  
25       to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM,  
30       INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the

CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:37-72. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative nucleic acid-associated proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode nucleic acid-associated proteins, the encoded polypeptides were analyzed by querying against PFAM models for nucleic acid-associated proteins. Potential nucleic acid-associated proteins were also identified by homology to Incyte cDNA sequences that had been annotated as nucleic acid-associated proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length

polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

5     **V.     Assembly of Genomic Sequence Data with cDNA Sequence Data**  
      **“Stitched” Sequences**

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then “stitched” together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbprl public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

20     **“Stretched” Sequences**

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions

may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

## VI. Chromosomal Mapping of NAAP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:37-72 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:37-72 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

## VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook et al., *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16).

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer

search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding NAAP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding NAAP.

### VIII. Extension of NAAP Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels,

fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

## **IX. Identification of Single Nucleotide Polymorphisms in NAAP Encoding**

### **Polynucleotides**

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:37-72 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins

or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

#### **X. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:37-72 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### **XI. Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned

technologies should be uniform and solid with a non-porous surface (Schena (1999), *supra*).

Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is

then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

### Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

### Hybridization

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide  
5 containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477,  
10 Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a  
15 cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore,  
20 are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC  
25 computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot  
30 is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements

that exhibited at least about a two-fold change in expression, a signal-to-background ratio of at least 2.5, and an element spot size of at least 40% were identified as differentially expressed using the GEMTOOLS program (Incyte Genomics).

### Expression

5 In one example, SEQ ID NO:41 showed differential expression in several human breast cancer cell lines, as determined by microarray analysis. HMEC is a human primary mammary epithelial cell strain derived from normal mammary tissue (Clonetics, San Diego, CA). The following cell lines were tested on microarrays: MCF-10A is a human breast mammary gland cell line isolated from a 36-year-old female with fibrocystic breast disease; SkBR3 is a breast adenocarcinoma cell line  
10 isolated from a malignant pleural effusion of a 43-year-old female; MCF7 is a breast adenocarcinoma cell line derived from the pleural effusion of a 69-year-old female; T47D is a breast carcinoma cell line derived from a pleural effusion from a 54-year-old female with an infiltrating ductal carcinoma of the breast; BT20 is a breast carcinoma cell line derived *in vitro* from cells emigrating out of thin slices of a tumor mass isolated from a 74-year-old female; MDA-mb-231 is a metastatic breast tumor cell  
15 line derived from the pleural effusion of a 51-year-old female with metastatic breast carcinoma. All cell cultures were propagated in media according to the supplier's recommendations and grown to 70-80% confluence prior to RNA isolation.

The expression of cDNAs from the five tumor cell lines representing various stages of breast tumor progression (BT20, MCF7, MDA-mb-231, SKBr3, and T47D) were compared with that of the  
20 non-malignant mammary epithelial cell lines, HMEC or MCF-10A.

SEQ ID NO:41 showed at least two-fold differential expression when comparing HMEC cells versus Sk-BR-3 and T-47D cells. Additionally, SEQ ID NO:41 expression was decreased at least two-fold when comparing breast cells from fibrocystic breast tissue versus BT-20, MCF-7, MDA-mb-231, Sk-BR-3, and T-47D cancerous cell lines. These experiments indicate that SEQ ID NO:41 was  
25 significantly under-expressed in the breast tumor cell lines tested, further establishing the utility of SEQ ID NO:41 as a diagnostic marker or as a potential therapeutic target for breast cancer.

In another example, SEQ ID NO:44 is upregulated 3.9 fold in DC as compared to monocytes, suggesting that SEQ ID NO:44, encoding SEQ ID NO:8, could be used for example, to understand the process by which monocytes differentiate into immature dendritic cells and eventually allow  
30 manipulation of the immune system leading to potential immunotherapies for diseases such as cancer, AIDS, and infectious diseases; and enhancing vaccine efficacy.

In another example, the expression of SEQ ID NO:48 is upregulated in six out of seven PBMC populations (each of which was obtained from a different donor) treated with Staphylococcal

exotoxins (SEB). The PBMCs were stimulated *in vitro* with SEB for 24 and 72 hours. The expression of SEQ ID NO:48 was higher after 24 hours and dropped after 72 hours. Therefore, SEQ ID NO:48 is useful in diagnostic assays for immune responses.

In yet another example, SEQ ID NO:64 showed differential expression in the C3A cell line, a well-established *in vitro* model of the mature human liver (Mickelson, J.K. et al. (1995) Hepatology 22:866-875; Nagendra, A.R. et al. (1997) Am. J. Physiol. 272:G408-G416), as determined by microarray analysis. The effects upon liver metabolism and hormone clearance mechanisms are important to understand the pharmacodynamics of a drug. For example, the human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was selected for strong contact inhibition of growth. The use of a clonal population enhances the reproducibility of the cells. C3A cells have many characteristics of primary human hepatocytes in culture: i) expression of insulin receptor and insulin-like growth factor II receptor; ii) secretion of a high ratio of serum albumin compared with  $\alpha$ -fetoprotein; iii) conversion of ammonia to urea and glutamine; iv) ability to metabolize aromatic amino acids; and v) proliferation in glucose-free and insulin-free medium. SEQ ID NO:64 showed differential expression in C3A cells treated with a variety of steroids including beclomethasone, medroxyprogesterone, budesonide, prednisone, dexamethasone, and progesterone, versus untreated C3A cells, as determined by microarray analysis. Therefore, SEQ ID NO:64 is useful for the diagnosis and monitoring of liver, endocrine, and reproductive diseases and in the diagnosis of and as a therapeutic target for inflammatory diseases and humoral immune response.

## **XII. Complementary Polynucleotides**

Sequences complementary to the NAAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring NAAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of NAAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the NAAP-encoding transcript.

## **XIII. Expression of NAAP**

Expression and purification of NAAP is achieved using bacterial or virus-based expression systems. For expression of NAAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA

transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).

Antibiotic resistant bacteria express NAAP upon induction with isopropyl beta-D-

5 thiogalactopyranoside (IPTG). Expression of NAAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding NAAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong  
10 polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

15 In most expression systems, NAAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences).  
20 Following purification, the GST moiety can be proteolytically cleaved from NAAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16).  
25 Purified NAAP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

#### XIV. Functional Assays

NAAP function is assessed by expressing the sequences encoding NAAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression  
30 vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either

liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of NAAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding NAAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding NAAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### XV. Production of NAAP Specific Antibodies

NAAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the NAAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel, 1995, *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A

peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel, 1995, *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-NAAP activity by, for example, binding the peptide or NAAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### **XVI. Purification of Naturally Occurring NAAP Using Specific Antibodies**

Naturally occurring or recombinant NAAP is substantially purified by immunoaffinity chromatography using antibodies specific for NAAP. An immunoaffinity column is constructed by covalently coupling anti-NAAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing NAAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of NAAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/NAAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and NAAP is collected.

#### **XVII. Identification of Molecules Which Interact with NAAP**

NAAP, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled NAAP, washed, and any wells with labeled NAAP complex are assayed. Data obtained using different concentrations of NAAP are used to calculate values for the number, affinity, and association of NAAP with the candidate molecules.

Alternatively, molecules interacting with NAAP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

NAAP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### **XVIII. Demonstration of NAAP Activity**

NAAP activity is measured by its ability to stimulate transcription of a reporter gene (Liu,

H.Y. et al. (1997) EMBO J. 16:5289-5298). The assay entails the use of a well characterized reporter gene construct, LexA<sub>op</sub>-LacZ, that consists of LexA DNA transcriptional control elements (LexA<sub>op</sub>) fused to sequences encoding the *E. coli* LacZ enzyme. The methods for constructing and expressing fusion genes, introducing them into cells, and measuring LacZ enzyme activity, are well known to those skilled in the art. Sequences encoding NAAP are cloned into a plasmid that directs the synthesis of a fusion protein, LexA-NAAP, consisting of NAAP and a DNA binding domain derived from the LexA transcription factor. The resulting plasmid, encoding a LexA-NAAP fusion protein, is introduced into yeast cells along with a plasmid containing the LexA<sub>op</sub>-LacZ reporter gene. The amount of LacZ enzyme activity associated with LexA-NAAP transfected cells, relative to control cells, is proportional to the amount of transcription stimulated by the NAAP.

Alternatively, NAAP activity is measured by its ability to bind zinc. A 5-10  $\mu$ M sample solution in 2.5 mM ammonium acetate solution at pH 7.4 is combined with 0.05 M zinc sulfate solution (Aldrich, Milwaukee WI) in the presence of 100  $\mu$ M dithiothreitol with 10% methanol added. The sample and zinc sulfate solutions are allowed to incubate for 20 minutes. The reaction solution is passed through a VYDAC column (Grace Vydac, Hesperia, CA) with approximately 300 Angstrom bore size and 5  $\mu$ M particle size to isolate zinc-sample complex from the solution, and into a mass spectrometer (PE Sciex, Ontario, Canada). Zinc bound to sample is quantified using the functional atomic mass of 63.5 Da observed by Whittall et al. (2000; Biochemistry 39:8406-8417).

In the alternative, a method to determine nucleic acid binding activity of NAAP involves a polyacrylamide gel mobility-shift assay. In preparation for this assay, NAAP is expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector containing NAAP cDNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of NAAP. Extracts containing solubilized proteins can be prepared from cells expressing NAAP by methods well known in the art. Portions of the extract containing NAAP are added to [<sup>32</sup>P]-labeled RNA or DNA. Radioactive nucleic acid can be synthesized *in vitro* by techniques well known in the art. The mixtures are incubated at 25°C in the presence of RNase- and DNase-inhibitors under buffered conditions for 5-10 minutes. After incubation, the samples are analyzed by polyacrylamide gel electrophoresis followed by autoradiography. The presence of a band on the autoradiogram indicates the formation of a complex between NAAP and the radioactive transcript. A band of similar mobility will not be present in samples prepared using control extracts prepared from untransformed cells.

In the alternative, a method to determine methylase activity of NAAP measures transfer of radiolabeled methyl groups between a donor substrate and an acceptor substrate. Reaction mixtures

(50  $\mu$ l final volume) contain 15 mM HEPES, pH 7.9, 1.5 mM  $MgCl_2$ , 10 mM dithiothreitol, 3% polyvinylalcohol, 1.5  $\mu$ Ci [*methyl*- $^3H$ ]AdoMet (0.375  $\mu$ M AdoMet) (DuPont-NEN), 0.6  $\mu$ g NAAP, and acceptor substrate (e.g., 0.4  $\mu$ g [ $^{35}S$ ]RNA, or 6-mercaptapurine (6-MP) to 1 mM final concentration). Reaction mixtures are incubated at 30°C for 30 minutes, then 65°C for 5 minutes.

5           Analysis of [*methyl*- $^3H$ ]RNA is as follows: (1) 50  $\mu$ l of 2 x loading buffer (20 mM Tris-HCl, pH 7.6, 1 M LiCl, 1 mM EDTA, 1% sodium dodecyl sulphate (SDS)) and 50  $\mu$ l oligo d(T)-cellulose (10 mg/ml in 1 x loading buffer) are added to the reaction mixture, and incubated at ambient temperature with shaking for 30 minutes. (2) Reaction mixtures are transferred to a 96-well filtration plate attached to a vacuum apparatus. (3) Each sample is washed sequentially with three 2.4 ml  
10           aliquots of 1 x oligo d(T) loading buffer containing 0.5% SDS, 0.1% SDS, or no SDS. (4) RNA is eluted with 300  $\mu$ l of water into a 96-well collection plate, transferred to scintillation vials containing liquid scintillant, and radioactivity determined.

          Analysis of [*methyl*- $^3H$ ]6-MP is as follows: (1) 500  $\mu$ l 0.5 M borate buffer, pH 10.0, and then 2.5 ml of 20% (v/v) isoamyl alcohol in toluene are added to the reaction mixtures. (2) The samples  
15           are mixed by vigorous vortexing for ten seconds. (3) After centrifugation at 700g for 10 minutes, 1.5 ml of the organic phase is transferred to scintillation vials containing 0.5 ml absolute ethanol and liquid scintillant, and radioactivity determined. (4) Results are corrected for the extraction of 6-MP into the organic phase (approximately 41%). For both [*methyl*- $^3H$ ]RNA and [*methyl*- $^3H$ ]6-MP, NAAP activity is proportional to the measured radioactivity.

20           Alternatively, DNA repair activity of DNAME is measured as incorporation of [ $^{32}P$ ]dATP into a plasmid treated with a DNA damaging agent, such as cisplatin or ultraviolet irradiation, relative to a control, untreated plasmid DNA (Coudore, F. et al. (1997) FEBS Lett. 414:581-584). Cell extracts are purified from mammalian cell lines, *E. coli*, or *S. cerevisiae* having compromised endogenous repair activities due to mutations in repair enzymes. Cell extracts are prepared by  
25           hypotonic lysis of cells followed by centrifugation at 300,000 x g. Extracts are treated with 63% ammonium sulfate to minimize non-specific nuclease activity. The repair synthesis assay is performed in a 50  $\mu$ l reaction volume containing 200  $\mu$ g protein in cell extract, 300 ng damaged plasmid, 300 ng control plasmid, 4  $\mu$ M dATP, 20  $\mu$ M each dCTP, dTTP, and dGTP, 0.2  $\mu$ M [ $^{32}P$ ]dATP, 20 mM HEPES-KOH (pH 7.8), 2.5  $\mu$ g creatine phosphokinase, 7 mM  $MgCl_2$ , and 2 mM EGTA. Identical  
30           reactions are set up with and without purified DNAME. After a 3 h incubation at 30°C, reaction mixtures are treated with 200  $\mu$ g/ml proteinase K and 0.5% SDS. Plasmid DNA is purified from reaction mixtures by phenol-chloroform extraction and ethanol precipitation. Data is quantified by gel electrophoresis of linearized plasmid followed by autoradiography, scintillation counting of excised

DNA bands, and densitometry of the photographic negative of the gel to normalize for plasmid DNA recovery.

In the alternative, type I topoisomerase activity of NAAP can be assayed based on the relaxation of a supercoiled DNA substrate. NAAP is incubated with its substrate in a buffer lacking  $Mg^{2+}$  and ATP, the reaction is terminated, and the products are loaded on an agarose gel. Altered topoisomers can be distinguished from supercoiled substrate electrophoretically. This assay is specific for type I topoisomerase activity because  $Mg^{2+}$  and ATP are necessary cofactors for type II topoisomerases.

Type II topoisomerase activity of NAAP can be assayed based on the decatenation of a kinetoplast DNA (KDNA) substrate. NAAP is incubated with KDNA, the reaction is terminated, and the products are loaded on an agarose gel. Monomeric circular KDNA can be distinguished from catenated KDNA electrophoretically. Kits for measuring type I and type II topoisomerase activities are available commercially from Topogen (Columbus OH).

ATP-dependent RNA helicase unwinding activity of NAAP can be measured by the method described by Zhang and Grosse (1994; Biochemistry 33:3906-3912). The substrate for RNA unwinding consists of  $^{32}P$ -labeled RNA composed of two RNA strands of 194 and 130 nucleotides in length containing a duplex region of 17 base-pairs. The RNA substrate is incubated together with ATP,  $Mg^{2+}$ , and varying amounts of NAAP in a Tris-HCl buffer, pH 7.5, at 37°C for 30 minutes. The single-stranded RNA product is then separated from the double-stranded RNA substrate by electrophoresis through a 10% SDS-polyacrylamide gel, and quantitated by autoradiography. The amount of single-stranded RNA recovered is proportional to the amount of NAAP in the preparation.

In the alternative, NAAP function is assessed by expressing the sequences encoding NAAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen Corporation, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu g$  of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2  $\mu g$  of an additional plasmid containing sequences encoding a marker protein are co-transfected.

Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; CLONTECH), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated laser optics-based technique, is

used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties.

FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of NAAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding NAAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Inc., Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding NAAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

Pseudouridine synthase activity of NAAP is assayed using a tritium ( $^3\text{H}$ ) release assay modified from Nurse et al. (1995; RNA 1:102-112), which measures the release of  $^3\text{H}$  from the  $\text{C}_5$  position of the pyrimidine component of uridylyate (U) when  $^3\text{H}$ -radiolabeled U in RNA is isomerized to pseudouridine ( $\psi$ ). A typical 500  $\mu\text{l}$  assay mixture contains 50 mM HEPES buffer (pH 7.5), 100 mM ammonium acetate, 5 mM dithiothreitol, 1 mM EDTA, 30 units RNase inhibitor, and 0.1-4.2  $\mu\text{M}$  [ $5\text{-}^3\text{H}$ ]tRNA (approximately 1  $\mu\text{Ci/nmol}$  tRNA). The reaction is initiated by the addition of  $<5\text{ }\mu\text{l}$  of a concentrated solution of NAAP (or sample containing NAAP) and incubated for 5 min at 37  $^\circ\text{C}$ . Portions of the reaction mixture are removed at various times (up to 30 min) following the addition of NAAP and quenched by dilution into 1 ml 0.1 M HCl containing Norit-SA3 (12% w/v). The quenched reaction mixtures are centrifuged for 5 min at maximum speed in a microcentrifuge, and the supernatants are filtered through a plug of glass wool. The pellet is washed twice by resuspension in 1 ml 0.1 M HCl, followed by centrifugation. The supernatants from the washes are separately passed through the glass wool plug and combined with the original filtrate. A portion of the combined filtrate is mixed with scintillation fluid (up to 10 ml) and counted using a scintillation counter. The amount of

$^3\text{H}$  released from the RNA and present in the soluble filtrate is proportional to the amount of pseudouridine synthase activity in the sample (Ramamurthy, V. (1999) J. Biol. Chem. 274:22225-22230).

In the alternative, pseudouridine synthase activity of NAAP is assayed at 30 °C to 37 °C in a mixture containing 100 mM Tris-HCl (pH 8.0), 100 mM ammonium acetate, 5 mM  $\text{MgCl}_2$ , 2 mM dithiothreitol, 0.1 mM EDTA, and 1-2 fmol of [ $^{32}\text{P}$ ]-radiolabeled runoff transcripts (generated *in vitro* by an appropriate RNA polymerase, i.e., T7 or SP6) as substrates. NAAP is added to initiate the reaction or omitted from the reaction in control samples. Following incubation, the RNA is extracted with phenol-chloroform, precipitated in ethanol, and hydrolyzed completely to 3-nucleotide monophosphates using RNase T<sub>2</sub>. The hydrolysates are analyzed by two-dimensional thin layer chromatography, and the amount of  $^{32}\text{P}$  radiolabel present in the  $\psi\text{MP}$  and UMP spots are evaluated after exposing the thin layer chromatography plates to film or a PhosphorImager screen. Taking into account the relative number of uridylate residues in the substrate RNA, the relative amount  $\psi\text{MP}$  and UMP are determined and used to calculate the relative amount of  $\psi$  per tRNA molecule (expressed in mol  $\psi$  /mol of tRNA or mol  $\psi$  /mol of tRNA/minute), which corresponds to the amount of pseudouridine synthase activity in the NAAP sample (Lecointe, *supra*).

$\text{N}^2, \text{N}^2$ -dimethylguanosine transferase (( $\text{m}^2_2\text{G}$ )methyltransferase) activity of NAAP is measured in a 160  $\mu\text{l}$  reaction mixture containing 100 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 10 mM  $\text{MgCl}_2$ , 20 mM  $\text{NH}_4\text{Cl}$ , 1mM dithiothreitol, 6.2  $\mu\text{M}$  *S*-adenosyl-L-[methyl- $^3\text{H}$ ]methionine (30-70 Ci/mM), 8  $\mu\text{g}$   $\text{m}^2_2\text{G}$ -deficient tRNA or wild type tRNA from yeast, and approximately 100  $\mu\text{g}$  of purified NAAP or a sample comprising NAAP. The reactions are incubated at 30 °C for 90 min and chilled on ice. A portion of each reaction is diluted to 1 ml in water containing 100  $\mu\text{g}$  BSA. 1 ml of 2 M HCl is added to each sample and the acid insoluble products are allowed to precipitate on ice for 20 min before being collected by filtration through glass fiber filters. The collected material is washed several times with HCl and quantitated using a liquid scintillation counter. The amount of  $^3\text{H}$  incorporated into the  $\text{m}^2_2\text{G}$ -deficient, acid-insoluble tRNAs is proportional to the amount of  $\text{N}^2, \text{N}^2$ -dimethylguanosine transferase activity in the NAAP sample. Reactions comprising no substrate tRNAs, or wild-type tRNAs that have already been modified, serve as control reactions which should not yield acid-insoluble  $^3\text{H}$ -labeled products.

Polyadenylation activity of NAAP is measured using an *in vitro* polyadenylation reaction. The reaction mixture is assembled on ice and comprises 10  $\mu\text{l}$  of 5 mM dithiothreitol, 0.025% (v/v) NONIDET P-40, 50 mM creatine phosphate, 6.5% (w/v) polyvinyl alcohol, 0.5 unit/ $\mu\text{l}$  RNAGUARD (Pharmacia), 0.025  $\mu\text{g}/\mu\text{l}$  creatine kinase, 1.25 mM cordycepin 5'-triphosphate, and 3.75 mM  $\text{MgCl}_2$ , in

a total volume of 25  $\mu$ l. 60 fmol of CstF, 50 fmol of CPSF, 240 fmol of PAP, 4  $\mu$ l of crude or partially purified CF II and various amounts of amounts CF I are then added to the reaction mix. The volume is adjusted to 23.5  $\mu$ l with a buffer containing 50 mM TrisHCl, pH 7.9, 10% (v/v) glycerol, and 0.1 mM Na-EDTA. The final ammonium sulfate concentration should be below 20 mM. The reaction is initiated (on ice) by the addition of 15 fmol of  $^{32}$ P-labeled pre-mRNA template, along with 2.5  $\mu$ g of unlabeled tRNA, in 1.5  $\mu$ l of water. Reactions are then incubated at 30 °C for 75-90 min and stopped by the addition of 75  $\mu$ l (approximately two-volumes) of proteinase K mix (0.2 M Tris-HCl, pH 7.9, 300 mM NaCl, 25 mM Na-EDTA, 2% (w/v) SDS), 1  $\mu$ l of 10 mg/ml proteinase K, 0.25  $\mu$ l of 20 mg/ml glycogen, and 23.75  $\mu$ l of water). Following incubation, the RNA is precipitated with ethanol and analyzed on a 6% (w/v) polyacrylamide, 8.3 M urea sequencing gel. The dried gel is developed by autoradiography or using a phosphoimager. Cleavage activity is determined by comparing the amount of cleavage product to the amount of pre-mRNA template. The omission of any of the polypeptide components of the reaction and substitution of NAAP is useful for identifying the specific biological function of NAAP in pre-mRNA polyadenylation (Rüegsegger, *supra*; and references within).

tRNA synthetase activity is measured as the aminoacylation of a substrate tRNA in the presence of [ $^{14}$ C]-labeled amino acid. NAAP is incubated with [ $^{14}$ C]-labeled amino acid and the appropriate cognate tRNA (for example, [ $^{14}$ C]alanine and tRNA<sup>ala</sup>) in a buffered solution.  $^{14}$ C-labeled product is separated from free [ $^{14}$ C]amino acid by chromatography, and the incorporated  $^{14}$ C is quantified by scintillation counter. The amount of  $^{14}$ C-labeled product detected is proportional to the activity of NAAP in this assay.

In the alternative, NAAP activity is measured by incubating a sample containing NAAP in a solution containing 1 mM ATP, 5 mM Hepes-KOH (pH 7.0), 2.5 mM KCl, 1.5 mM magnesium chloride, and 0.5 mM DTT along with misacylated [ $^{14}$ C]-Glu-tRNA<sup>Gln</sup> (e.g., 1  $\mu$ M) and a similar concentration of unlabeled L-glutamine. Following the quenching of the reaction with 3 M sodium acetate (pH 5.0), the mixture is extracted with an equal volume of water-saturated phenol, and the aqueous and organic phases are separated by centrifugation at 15,000  $\times$  g at room temperature for 1 min. The aqueous phase is removed and precipitated with 3 volumes of ethanol at -70°C for 15 min. The precipitated aminoacyl-tRNAs are recovered by centrifugation at 15,000  $\times$  g at 4°C for 15 min. The pellet is resuspended in 25 mM KOH, deacylated at 65°C for 10 min., neutralized with 0.1 M HCl (to final pH 6-7), and dried under vacuum. The dried pellet is resuspended in water and spotted onto a cellulose TLC plate. The plate is developed in either isopropanol/formic acid/water or ammonia/water/chloroform/ methanol. The image is subjected to densitometric analysis and the relative amounts of Glu and Gln are calculated based on the R<sub>f</sub> values and relative intensities of the

spots. NAAP activity is calculated based on the amount of Gln resulting from the transformation of Glu while acylated as Glu-tRNA<sup>Gln</sup> (adapted from Curnow, A.W. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11819-26).

#### **XIX. Identification of NAAP Agonists and Antagonists**

5 Agonists or antagonists of NAAP activation or inhibition may be tested using the assays described in section XVIII. Agonists cause an increase in NAAP activity and antagonists cause a decrease in NAAP activity.

10 Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be 15 understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be 20 defined by the following claims and their equivalents.

Table 1

IncYTE Project ID	Polypeptide SEQ ID NO:	IncYTE Polypeptide ID	Polynucleotide SEQ ID NO:	IncYTE Polynucleotide ID	CA2 Reagents
7490148	1	7490148CD1	37	7490148CB1	
7490301	2	7490301CD1	38	7490301CB1	
2383223	3	2383223CD1	39	2383223CB1	90088564CA2, 90088632CA2
3495982	4	3495982CD1	40	3495982CB1	3495982CA2
7477891	5	7477891CD1	41	7477891CB1	
72688352	6	72688352CD1	42	72688352CB1	
7490652	7	7490652CD1	43	7490652CB1	
7489744	8	7489744CD1	44	7489744CB1	
3363382	9	3363382CD1	45	3363382CB1	
7491148	10	7491148CD1	46	7491148CB1	
8126343	11	8126343CD1	47	8126343CB1	
7044055	12	7044055CD1	48	7044055CB1	
7493424	13	7493424CD1	49	7493424CB1	
1482140	14	1482140CD1	50	1482140CB1	
394992	15	394992CD1	51	394992CB1	
5093550	16	5093550CD1	52	5093550CB1	
7487977	17	7487977CD1	53	7487977CB1	90175683CA2, 90175691CA2
1706514	18	1706514CD1	54	1706514CB1	90088114CA2
7488247	19	7488247CD1	55	7488247CB1	
1427269	20	1427269CD1	56	1427269CB1	
103135	21	103135CD1	57	103135CB1	2445787CA2
1907346	22	1907346CD1	58	1907346CB1	
3041036	23	3041036CD1	59	3041036CB1	
3856879	24	3856879CD1	60	3856879CB1	
4178665	25	4178665CD1	61	4178665CB1	90105075CA2
7493326	26	7493326CD1	62	7493326CB1	
1553836	27	1553836CD1	63	1553836CB1	90092916CA2, 90092932CA2, 90097748CA2, 90097824CA2
1908201	28	1908201CD1	64	1908201CB1	
2827615	29	2827615CD1	65	2827615CB1	
4304550	30	4304550CD1	66	4304550CB1	
7473738	31	7473738CD1	67	7473738CB1	
4447743	32	4447743CD1	68	4447743CB1	
7497554	33	7497554CD1	69	7497554CB1	90162016CA2
7475843	34	7475843CD1	70	7475843CB1	6030485CA2
6319550	35	6319550CD1	71	6319550CB1	
7510064	36	7510064CD1	72	7510064CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7490148	g18676392	1.0E-177	[fl][Homo sapiens] homeobox protein GSH-2 Hsieh-Li, H. M. et al. (1995) Gsh-2, a murine homeobox gene expressed in the developing brain. Mech. Dev. 50:177-186.
2	7490301	g2645175	1.0E-57	[Homo sapiens] TFIIID subunit p30beta
3	2383223	g6467206	6.6E-207	[Homo sapiens] gonadotropin inducible transcription repressor-4 Kuzuhara, T. and Horikoshi, M. (1996) Isolation and characterization of a cDNA encoding a human TFIIID subunit containing a variety of putative structural motifs including direct repeats. Biol. Pharm. Bull. 19:122-126.
4	3495982	g6467206	5.8E-112	[Homo sapiens] gonadotropin inducible transcription repressor-4 Kuzuhara and Horikoshi, <i>supra</i> .
5	7477891	g8176525	9.6E-190	[Homo sapiens] interferon-inducible myeloid differentiation transcriptional activator Trapani, J.A. et al. (1994) Genomic organization of IFI16, an interferon-inducible gene whose expression is associated with human myeloid cell differentiation: correlation of predicted protein domains with exon organization. Immunogenetics 40:415-424.
6	72688352	g10442700	1.4E-151	[Homo sapiens] zinc-finger protein ZBRK1 Zheng, L. et al. (2000) Sequence-specific transcriptional corepressor function for BRCA1 through a novel zinc finger protein, ZBRK1. Mol. Cell 6:757-768.
7	7490652	g854179	3.2E-95	[Homo sapiens] ribosomal protein S3a Nolte, D. et al. (1996) The human S3a ribosomal protein: sequence, location and cell-free transcription of the functional gene. Gene 169:179-185.
8	7489744	g181486	5.8E-199	[Homo sapiens] DNA-binding protein B Sakura, H. et al. (1988) Two human genes isolated by a novel method encode DNA-binding proteins containing a common region of homology. Gene 73:499-507.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
9	3363382	g2460119	8.2E-161	[Mus musculus] vascular actin single-stranded DNA-binding factor 2 p44 component; purine-rich single-stranded DNA-binding protein beta; PurB beta Kelm, J. Jr. et al. (1997) Sequence of cDNAs encoding components of vascular actin single-stranded DNA-binding factor 2 establish identity to Puralpha and Purbeta. J. Biol. Chem. 272:26727-26733.
10	7491148	g17046404	0.0	[f][Homo sapiens] RNA-binding protein CELF6
11	8126343	g20072019	1.0E-122	[Mus musculus] similar to integral inner nuclear membrane protein MAN1 Lin, F. et al. (2000) MAN1, an inner nuclear membrane protein that shares the LEM domain with lamina-associated polypeptide 2 and emerlin. J. Biol. Chem. 275:4840-4847.
12	7044055	g13194724	2.6E-292	[Homo sapiens] integral inner nuclear membrane protein MAN1
13	7493424	g3435244	0.0	[Homo sapiens] cytokine-like nuclear factor n-pac
14	1482140	g18448895	0.0	[Homo sapiens] centriole associated protein CEP110
		g8953575	0.0	[f][Homo sapiens] inhibitory PAS domain protein
			1.20E-262	[Rattus norvegicus] hypoxia inducible factor 3 alpha Kietzmann, T. et al. (2001) Perivenous expression of the mRNA of the three hypoxia-inducible factor alpha-subunits, HIF1alpha, HIF2alpha and HIF3alpha, in rat liver. Biochem. J. 354:531-537.
15	394992	g6683492	0.0	[Homo sapiens] bromodomain PHD finger transcription factor Jones, M.H. et al. (2000) Identification and characterization of BPTF, a novel bromodomain transcription factor. Genomics 63:35-39.
16	5093550	g1017722	4.4E-249	[Homo sapiens] repressor transcriptional factor
17	7487977	g8547213	1.5E-129	[Homo sapiens] HMG box transcription factor Hargrave, M. et al. (2000) Fine mapping of the neurally expressed gene SOX14 to human 3q23, relative to three congenital diseases. Hum. Genet. 106:432-439.
18	1706514	g6467206	7.3E-137	[Homo sapiens] gonadotropin inducible transcription repressor-4

Table 2

Polypeptide SEQ ID NO:	Incye Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
19	7488247	g5002200	2.0E-70	[Mus musculus] msg1-related protein 2 Yahata, T. et al. (2000) The MSG1 non-DNA-binding transactivator binds to the p300/CBP coactivators, enhancing their functional link to the Smad transcription factors. J. Biol. Chem 275:8825-8834.
20	1427269	g6016005	8.0E-140	[Homo sapiens] CoREST protein Andres, M.E. et al. (1999) CoREST: a functional corepressor required for regulation of neural-specific gene expression. Proc. Natl. Acad. Sci. USA 96:9873-9878.
21	103135	g8163824	6.7E-71	[Homo sapiens] krueppel-like zinc finger protein HZF2 Schafer, U. et al. (2000) Identification of a nitric oxide-regulated zinc finger containing transcription factor using motif-directed differential display. Biochim. Biophys. Acta 1494:269-276.
22	1907346	g14348588	0.0	[3' incm][Homo sapiens] KRAB zinc finger protein
23	3041036	g9502202	2.9E-284	[Homo sapiens] endothelial zinc finger protein induced by tumor necrosis factor alpha Mataki, C. et al. (2000) A novel zinc finger protein mRNA in human umbilical vein endothelial cells is profoundly induced by tumor necrosis factor alpha. J. Atheroscler. Thromb. 7:97-103.
24	3856879	g1020145	8.8E-156	[Homo sapiens] DNA binding protein Belletucci, L.J. et al. (1989) The human genome contains hundreds of genes coding for finger proteins of the Kruppel type. DNA 8:377-387.
25	4178665	g1769491	6.6E-159	[Homo sapiens] krueppel-related zinc finger protein Goldwurm, S. et al. (1997) Identification of a novel Krueppel-related zinc finger gene (ZNF184) mapping to 6p21.3. Genomics 40:486-489.
26	7493326	g1020145	9.3E-208	[Homo sapiens] DNA binding protein Bellefroid, <i>supra</i> .
29	2827615	g13560888	1.1E-174	[Homo sapiens] E2F1-related protein 1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
30	4304550	g6901197	2.9E-137	[Schizosaccharomyces pombe] putative helicase
31	7473738	g2245023	2.4E-133	[Arabidopsis thaliana] RNA helicase
32	4447743	g38032	5.8E-235	[Homo sapiens] ZNF43 Lovering, R. and Trowsdale, J. (1991) A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell lines. Nucleic Acids Res. 19:2921-2928.
33	7497554	g3169261	0.0	[fl][Mus musculus] T-box transcription factor Agulnik, S.I. et al. (1998) Cloning, mapping, and expression analysis of TBX15, a new member of the T-Box gene family. Genomics 51:68-75.
34	7475843	g1017722	7.9E-128	[Homo sapiens] repressor transcriptional factor
35	6319550	g3287501	4.9E-238	[Mus musculus] BAZF Okabe, S. et al. (1998) BAZF, a novel Bcl6 homolog, functions as a transcriptional repressor. Mol. Cell. Biol. 18:4235-4244.
36	7510064	g16040975	0.0	[Homo sapiens] (AB054067) hypoxia-inducible factor-3 alpha Hara, S. et al. (2001) Expression and characterization of hypoxia-inducible factor (HIF)-3alpha in human kidney: suppression of HIF-mediated gene expression by HIF-3alpha. Biochem. Biophys. Res. Commun. 287:808-813.
		629484 Hif3a	1.4E-267	[Rattus norvegicus] Hypoxia-inducible factor (HIF) alpha-subunit, activates expression from genes with HIF-responsive elements Kietzmann et al., <i>supra</i> .
		335776 HIF1A	5.9E-121	[Homo sapiens] [DNA-binding protein; Transcription factor] [Nuclear] Basic helix-loop-helix transcription factor that contains a PAS domain, heterodimerizes with the Ah receptor nuclear translocator (ARNT) and mediates transcriptional responses to hypoxia and dioxin-signaling Maxwell, P.H. et al. (1999) The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. Nature 399:271-275.

Table 3

SEQ ID NO:	Incyle Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7490148CD1	304	S21 S52 S71 S242 S284 S294 T15 T264 Y238	N240	Signal Peptide: M43-G74	HMMER
					Signal Peptide: M43-L69	HMMER
					Homeobox domain: K203-K259	HMMER_PFAM
					Transmembrane domain: G145-A162	TMAP
					Homeobox antennapedia-type protein BL00032: R206-K244, Q245-K262	BLIMPS_BLOCKS
					Homeobox domain protein BL00027: L217-K259	BLIMPS_BLOCKS
					Homeobox domain signature and profile: E216- V279	PROFILESAN
					Signal_cleavage: M37-G74	SPSCAN
					Homeotic antennapedia protein signature PR00025: K203-E218	BLIMPS_PRINTS
					HOMEBOX PROTEIN TRANSCRIPTION DNA-BINDING NUCLEAR DEVELOPMENTAL REGULATION GSH2 GSH1 ACTIVATOR PD027913: M1-S72	BLAST_PRODROM
					HOMEBOX PROTEIN GSH2 TRANSCRIPTION REGULATION DNA-BINDING NUCLEAR DEVELOPMENTAL PD048406: L163-G202 PD049345: G261-L304	BLAST_PRODROM
					PROTEIN HOMEBOX DNA-BINDING NUCLEAR DEVELOPMENTAL TRANSCRIPTION REGULATION FACTOR HOMEODOMAIN METAL-BINDING PD000010: K203-E260	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					HOMEBOX DM00009 P31316 199-262: Q198-K262 DM00009 P31315 144-205: P200-K262 DM00009 P50901 128-193: G202-E260 DM00009 S34164 172-238: P200-K262	BLAST_DOMO
					Homeobox domain signature: I235-K258	MOTIFS
2	7490301CD1	198	S23 S107 S133 T3 T55 T74 T85 T130		Transmembrane domain: S133-G151; N-terminus is cytosolic	TMAP
					PROTEIN TRANSCRIPTION FACTOR TFIIID INITIATION SUBUNIT REGULATION NUCLEAR BETA P28BETA PD013308: D87-P189	BLAST_PRODOM
					TRANSCRIPTION INITIATION FACTOR TFIIID 28 KD SUBUNIT TAFII28 P30BETA REGULATION NUCLEAR PROTEIN PD097951: E2-V86	BLAST_PRODOM
3	2383223CD1	576	S14 S24 S42 S202 S230 S338 S566 T36 T91 T123	N87 N405	KRAB box: V4-D54	HMMER_PFAM
					Zinc finger, C2H2 type: Y272-H294, Y132-H154, Y244-H266, Y160-H182, Y216-H238, Y188-H210, Y468-490, Y300-H322, Y100-H122, Y356-H378, L552-H574, Y440-H462, Y524-H546, Y412-H434, Y384-H406, Y496-H518, Y328-H350	HMMER_PFAM
					Zinc finger, C2H2 type BL00028: C358-H374	BLIMPS_BLOCKS
					C2H2-type zinc finger signature PR00048: P355-I368	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN ZINC FINGER ZINC	BLIMPS_PRODUM
					PD01066: F6-G44	
					PROTEIN ZINC-FINGER METAL-BINDING	BLIMPS_PRODUM
					PD00066: H234-C246	
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING ZINC FINGER PATERNALLY EXPRESSED ZN-FINGER PW1	BLAST_PRODUM
					PD017719: P131-F365, G324-H574, H66-I282	
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT	BLAST_PRODUM
					PD000072: K98-C165, K326-C389, K466-C529, K186-C249, K522-N575, R270-C333, K410-C473, K298-C361, K494-C557, K354-C417, K214-C277, P439-C501, K158-C221, K130-C193, K242-C305	
					HYPOTHETICAL ZINC FINGER PROTEIN B03B8.4 IN CHROMOSOME III ZINC FINGER DNA-BINDING METAL-BINDING NUCLEAR	BLAST_PRODUM
					PD149420: E185-Y496, T257-A560, E82-G240	
					ZINC FINGER PROTEIN 97 DNA-BINDING PROTEIN DNA-BINDING ZINC FINGER METAL-BINDING	BLAST_PRODUM
					PD071370: K135-K242	
					KRAB BOX DOMAIN	
					DM00605 P52737 1-76: M1-E76	BLAST_DOMO
					DM00605 I49636 10-85: D2-R64	

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 Q05481 789-829: E291-E332, E207-E248, R348-K388, E459-E500, E179-Q220, R516-Q556, E235-E276, E375-E416, E319-E360, K487-E528, H266-Q30, E266-Q3044 DM00002 Q05481 831-885: C277-E332, C193-E248, C305-E360	BLAST_DOMO
					Zn-finger, C2H2 type: C102-H122, C134-H154, C162-H182, C190-H210, C218-H238, C246-H266, C274-H294, C302-H322, C330-H350, C358-H378, C386-H406, C414-H434, C442-H462, C470-H490, C498-H518, C526-H546, C554-H574	MOTIFS
4	3495982CD1	426	S5 S49 S58 S177 S208 S236 S348 T52 T67 T89 T145 T248 Y189	N4 N65 N271	KRAB box: I57-Q103	HMMER_PFAM
					Zinc finger, C2H2 type: Y250-H272, H194-H216, Y222-H244, C156-H178, Y306-H328, Y278-H300, F334-H356	HMMER_PFAM
					Zinc finger, C2H2 type BL00028: C196-H212	BLIMPS_BLOCKS
					C2H2-type zinc finger signature PR00048: L321-G330, P193-S206	BLIMPS_PRINTS
					PROTEIN ZINC FINGER ZINC PD01066: L59-G97	BLIMPS_PRODROM
					PROTEIN ZINC FINGER METAL-BINDING PD00066: H296-C308	BLIMPS_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING ZINC FINGER PATERNALLY EXPRESSED ZN-FINGER PW1 PD017719: G152-R358	BLAST_PRODROM
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K154-C227, K220-C283, K276-C339, K192-C255, P249-C311	BLAST_PRODROM
					ZINC FINGER PROTEIN 101 ZINC FINGER METAL-BINDING DNA-BINDING PD098683: E98-K276	BLAST_PRODROM
					KRAB BOX DOMAIN DM00605 P52737 1-76: D55-E129 DM00605 I49636 10-85: D55-R117	BLAST_DOMO
					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 Q0548 789-829: R242-Q282, G190-E226, E297-E338, E213-E254 DM00002 P08042 314-358: C199-H244	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C156-H178, C158-H178, C196-H216, C224-H244, C252-H272, C280-H300, C308-H328, C336-H356	MOTIFS
5	747789ICD1	786	S95 S113 S134 S181 S185 S222 S266 S270 S279 S569 S776 T74 T81 T117 T140 T333 T421 T443 T499 T500 T556 T563 T607 T626 T673 T704 T712 T737 T764 T780	N179 N212 N264 N375 N585	HIN-200/IF120x domain: Q575-A742, F201-R371	HMMER_PFAM

Table 3

SEQ ID NO:	IncYTE Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PAAD/DAPIN DOMAIN: N4-T98	HMMER_PFAM
					PROTEIN INTERFERON INDUCTION NUCLEAR REPEAT INTERFERON ACTIVATABLE INTERFERON INDUCIBLE MYELOID DIFFERENTIATION ACTIVATOR PD007764: M513-K766, Q128-K391	BLAST_PRODROM
					INTERFERON-ACTIVATABLE PROTEIN 204 IFI204 INTERFERON INDUCTION NUCLEAR REPEAT PD134308: T117-F386, T541-L741	BLAST_PRODROM
					PROTEIN INTERFERON INDUCTION NUCLEAR REPEAT INTERFERON-ACTIVATABLE MYELOID DIFFERENTIATION ACTIVATOR DNA-BINDING PD014209: Y5-K90	BLAST_PRODROM
					GAMMA-INTERFERON-INDUCIBLE PROTEIN IFI16 INTERFERON-INDUCIBLE MYELOID DIFFERENTIATION TRANSCRIPTIONAL ACTIVATOR INTERFERON INDUCTION NUCLEAR REPEAT DNA-BINDING TRANSCRIPTION REGULATION PD134282: N394-L462, N454-L518	BLAST_PRODROM
					do NUCLEAR; DIFFERENTIATION; MYELOID; ANTIGEN; DM02433 P41218 166-397: T160-K391, A535-K766 DM02433 P15091 15-223: G165-R371, A535-L741 DM02433 P15092 205-416: T162-R371, T541-L741 DM02433 P15092 418-639: P565-K766, P205-I389	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	72688352CD1	617	S52 S91 S242 S326 S438 S466 S543 T9 T18 T263 T347 T451 T485	N40 N139 N253	Zinc finger, C2H2 type: Y260-H282, Y484-H506, Y372-H394, F456-H478, H316-H338, H204-H226, Y344-H366, Y400-H422, H232-H254, Y288-H310, Y428-H450, Y512-H535 KRAB box: L8-E69	HMMER_PFAM
					Zinc finger, C2H2 type BL00028: C234-H250	HMMER_PFAM BLIMPS_BLOCKS
					C2H2-type zinc finger signature PR00048: P427-K440	BLIMPS_PRINTS
					PROTEIN ZINC FINGER ZINC PD01066: L10-G48	BLIMPS_PRODROM
					PROTEIN ZINC FINGER METAL-BINDING PD00066: H334-C346	BLIMPS_PRODROM
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING ZINC FINGER PATERNALLY EXPRESSED ZN-FINGER PW1 PD017719: G256-G513, P203-H450, G284-F521, G228-S464, E180-I410, G312-L527, G340-E538	BLAST_PRODROM
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: L8-E69	BLAST_PRODROM
					ZINC FINGER PROTEIN 84 HPF2 ZINC FINGER DNA-BINDING METAL-BINDING NUCLEAR REPEAT PD097306: E69-H226	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K202-C265, K286-C349, P455-C517, K314-C377, K230-C293	BLAST_PRODOM
					KRAB BOX DOMAIN DM00605 I48689 1-85: Q5-P78 DM00605 P51523 5-79: Q5-P78 DM00605 P52736 1-72: L8-P78 DM00605 P52738 3-77: Q5-I72	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): G156-S163	MOTIFS
					Zn finger, C2H2 type: C206-H226, C234-H254, C262-H282, C290-H310, C318-H338, C346-H366, C374-H394, C402-H422, C430-H450, C458-H478, C486-H506, C514-H535	MOTIFS
7	7490652CDI	249	S26 S61 S70 S223 T115 T138 T185 T227	N221	Ribosomal S3Ae family: G12-K212	HMMER_PFAM
					Ribosomal protein S3Ae proteins BL01191: G13-K63, K85-C135, A170-N221	BLIMPS_BLOCKS
					PROTEIN RIBOSOMAL 40S S3A S3AE 30S CYC07 V-FOS TRANSFORMATION EFFECTOR PD003035: G13-K212	BLAST_PRODOM
					do RIBOSOMAL; S3A DM01430 JC4662 1-259: M1-E242 DM01430 P49395 1-259: A2-E242 DM01430 P49397 1-252: A2-L244 DM01430 S36622 1-254: M1-E236	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	7489744CD1	384	S63 S81 S162 S236 S374 T42 T122 T140 Y205	N138 N184 N372	signal_cleavage: M1-T67	SPSCAN
					Cold-shock DNA-binding domain: K118-P188	HMMER_PFAM
					Cold-shock DNA-binding BL00352: G121-I135, H147-V185	BLIMPS_BLOCKS
					Cold-shock DNA-binding domain signature: G101-V173	PROFILESAN
					Cold shock protein signature PR00050: G121-N136, E142-I151, R161-G179	BLIMPS_PRINTS
					TRANSCRIPTION DNA-BINDING REGULATION PROTEIN BINDING NUCLEAR REPRESSOR Y- BOX FACTOR BOX PD003149: G233-E384, G187-Q338	BLAST_PRODROM
					DNA-BINDING TRANSCRIPTION REGULATION PROTEIN BINDING BOX Y PROTEIN 1 Y-BOX NUCLEAR PD004557: G187-N270	BLAST_PRODROM
					TRANSCRIPTION FACTOR EF1 PD054259: D84-V123	BLAST_PRODROM
					DNA-BINDING TRANSCRIPTION REGULATION PROTEIN BINDING BOX Y PROTEIN 1 Y-BOX NUCLEAR PD149839: L102-G131	BLAST_PRODROM
					COLD-SHOCK DNA-BINDING DOMAIN DM02820 P16990 129-235: G189-N296 DM03030 P16990 237-323: M297-E384 DM00679 P16990 55-127: I115-P188 DM00679 P16989 87-159: I115-P188	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Cold-shock DNA-binding domain signature: Y132-I151	MOTIFS
9	3363382CD1	312	S6 S8 S36 S86 S100 S229 S253 S304 Y291		PUR-ALPHA DNA-BINDING PROTEIN TRANSCRIPTIONAL ACTIVATOR PURINE-RICH SINGLE-STRANDED ALPHA TRANSCRIPTION PD011221: P115-E249	BLAST_PRODROM
					PUR-ALPHA DNA-BINDING TRANSCRIPTIONAL ACTIVATOR PROTEIN PURINE-RICH SINGLE-STRANDED ALPHA TRANSCRIPTION PD013475: E28-A106	BLAST_PRODROM
					DNA-BINDING TRANSCRIPTIONAL ACTIVATOR PROTEIN PUR-ALPHA PURINE- RICH SINGLE-STRANDED ALPHA TRANSCRIPTION PD022742: V250-K289	BLAST_PRODROM
10	7491148CD1	441	S142 S384 T131 T173 T282	N394	RNA recognition motif (a.k.a. RRM, RBD, or RNP domain): L358-V429, L95-L166, L9-N81 Transmembrane domain: G226-A253, I360-F388; N-terminus is cytosolic Eukaryotic RNA-binding region RNP-1 proteins BL00030: L9-F27, K397-N406 Paraneoplastic encephalomyelitis antigen family signature (has RNA binding domain) PR00961: C356-E371, A413-K428 RIBONUCLEOPROTEIN NUCLEOPROTEIN PD030742: M182-F338	HMMER_PFAM  TMAP  BLIMPS_BLOCKS  BLIMPS_PRINTS  BLAST_PRODROM

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					RIBONUCLEOPROTEIN REPEAT DM00012 P28659 I-59: L112-A171 DM00012 P16914 397-479: G352-K432	BLAST_DOMO
11	8126343CD1	493	S44 S121 S173 S177 S178 S356 S369 T65 T387	N350	Transmembrane Domain: G242-V270, R407-L435	TMAP
12	7044055CD1	553	S5 S104 S113 S122 S130 S148 S152 S167 S223 S322 S361 S540 T182 T188 T205 T225 T236 T264 T266 T316 T367 T450 T467	N103 N112 N260 N297	PWWP domain: S5-R78	HMMER_PFAM
					Transmembrane Domain: D267-V295, F426-G454; N-terminus is cytosolic	TMAP

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	7493424CD1	1726	S6 S24 S150 S279 S287 S302 S373 S434 S474 S580 S599 S668 S672 S676 S774 S785 S796 S800 S858 S881 S923 S985 S1095 S1112 S1172 S1228 S1265 S1270 S1316 S1410 S1435 S1476 S1498 S1523 S1533 S1572 S1603 S1677 S1703 S1708 S1724 T75 T102 T145 T181 T329 T485 T549 T851 T921 T1084 T1254 T1338 T1414 T1418 T1453 T1471 T1512 T1560 T1611 T1671 T1692	N139 N529 N719 N748 N1622 N1722	IB3/SPOLYPEPTIDE PD131523: Q504-P755 PD098406: N245-A337, R335-R356	BLAST_PRODOM
					HEPTAD REPEAT PATTERN REPEAT (nuclear) DM05319 P30427 568-1938: V846-E1365, S800-L1608, I14-R588, E772-E969	BLAST_DOMO
					myosin-like protein MLP1 DM07884 P40457 16-1657: E770-A1697, K5-L528	BLAST_DOMO
					Leucine zipper pattern: L807-L828, L876-L897, L1275-L1296, L1468-L1489	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	1482140CD1	747	S46 S54 S69 S163 S171 S195 S208 S308 S316 S385 S403 S413 S439 S508 S538 S598 S619 S620 S665 T203 T216 T228 T295 T517 T615 Y307	N155 N233 N546	PAS domain: N127-Q191, K293-L336	HMMER_PFAM
					Myc-type, 'helix-loop-helix' dimerization domain proteins: BL00038: A67-A82, S92-R112	BLIMPS_BLOCKS
					Nuclear translocator signature PR00785: A93-L113, V142-G161, D174-Q192, W229-S248, A283-R300	BLIMPS_PRINTS
					PROTEIN NUCLEAR REPEAT TRANSCRIPTION REGULATION DNABINDING RECEPTOR HYDROCARBON ARYL BIOLOGICAL PD008299: E204-V387	BLAST_PRODROM
					PROTEIN TRANSCRIPTION REGULATION DNABINDING REPEAT NUCLEAR RECEPTOR HYDROCARBON ARYL DEVELOPMENTAL PD002640: K60-G117	BLAST_PRODROM
					do AH; MINDED; SINGLE; HYPOXIA; DM03763 38972 18-524: R59-E400 DM03763 P05709 1-564: E204-V497, K60-T189 DM03763 P35869 28-512: C114-T203, E204-R370, K62-R109 DM03763 P30561 27-503: K62-R370, P372-G391	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	394992CD1	2759	S3 S19 S54 S93 S197 S327 S356 S405 S420 S433 S511 S551 S554 S571 S615 S623 S726 S873 S879 S881 S891 S929 S931 S958 S974 S981 S982 S1022 S1043 S1065 S1095 S1146 S1172 S1184 S1211 S1221 S1225 S1226 S1238 S1286 S1319 S1413 S1415 S1430 S1445 S1475 S1503 S1539 S1572 S1607 S1631 S1639 S1640 S1682 S1714 S1736 S2244 S2489 S2501 S2504 S2577 S2615 S2630 T186 T344 T399 T437 T444 T471 T506 T624 T703 T709 T777 T916 T979 T996 T1081 T1309 T1380 T1469 T1474 T1507 T1584 T1690 T1697 T1730 T1731 T1741 T1774 T1798 T1891 T1958 T2219 T2339 T2513 T2517 T2521 T2529 T2555 T2587 T2642 T2689 Y646 Y1680 Y1754 Y2718	N163 N176 N406 N442 N446 N501 N504 N690 N964 N1021 N1025 N1210 N1215 N1236 N1307 N1403 N1480 N1962 N2127 N2388	PHD-finger: H253-H298, Y2524-A2573, Y2582-T2631	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Bromodomain: Y2646-V2732	HMMER_PFAM
					Transmembrane Domain: M737-R765;	TMAP
					N-terminus is non-cytosolic	
					Bromodomain proteins	BLIMPS_BLOCKS
					BL00633: L2656-P2668, P2675-Y2699, D2708-N2720	
					Bromodomain signature and profile: P2668-K2740	PROFLESCAN
					Wilm's tumour protein signature	BLIMPS_PRINTS
					PR00049: A2460-V2474	
					Bromodomain signature	BLIMPS_PRINTS
					PR00503: K2660-D2673, A2674-M2690, M2690-D2708, D2708-Y2727	
					FETAL ALZHEIMER ANTIGEN ALZ50	BLAST_PRODOM
					REACTIVE CLONE 1 NUCLEAR PROTEIN	
					PD143888: D369-W724	
					PROTEIN FETAL ALZHEIMER ANTIGEN ALZ50	BLAST_PRODOM
					REACTIVE CLONE NUCLEAR F26H11.2	
					PD037519: P78-S197	
					PD043262: I317-R403	
					BROMODOMAIN	BLAST_DOMO
					DM00265 Q03330 324-432: L2649-K2742	
					DM00265 S55259 894-1006: D2633-K2742	
					DM00265 P13709 27-141: L2649-K2742	
					DM00265 P25440 20-134: K2650-L2741	
					ATP/GTP-binding site motif A (P-loop): G1011-T1018	MOTIFS
					Bromodomain: A2662-Y2719	MOTIFS

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Cytochrome c heme-binding site: C254-K259	MOTIFS
					EGF-like domain: C2525-C2540, C2583-C2598	MOTIFS
16	5093550CD1	613	S8 S42 S84 S113 S154 S191 S247 S457 S611 T48 T159 T237 T265 T405 T433 T489 T601	N202 N230 N258 N286 N314 N318 N342 N370 N398 N458 N482 N510 N542 N566 N598	Zinc finger, C2H2 type: Y528-H550, Y556-H578, Y332-H354, Y360-H382, Y500-H522, Y416-H438, F164-H186, Y220-H242, Y192-H214, Y248-H270, Y444-H466, Y388-H410, Y304-H326, Y276-H298, Y584-H606, Y472-H494	HMMER_PFAM
					KRAB box: D2-E60	HMMER_PFAM
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN ZINC FINGER NUCLEAR TRANSCRIPTION REGULATION REPEAT PD008015: R61-K158	BLAST_PRODROM
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING ZINC FINGER PATERNALLY EXPRESSED ZN-FINGER PW1 PD017719: G272-H522	BLAST_PRODROM
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: D2-E60	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K470-C533, K330-C393, K358-C421, K414-C477, K386-C449, K498-C561, K302-C365, K442-C505, K526-C589, Y276-C337, K554-K610	BLAST_PRODUM
					KRAB BOX DOMAIN DM00605 P28160 1-69: M1-P68 DM00605 Q03923 1-75: D2-P68 DM00605 Q05481 10-83: D2-P68 DM00605 S22564 1-63: F7-P68	BLAST_DOMO
					Zn-finger, C2H2, type: C166-H186, C222-H242, C278-H298, C306-H326, C334-H354, C362-H382, C390-H410, C418-H438, C446-H466, C474-H494, C502-H522, C530-H550, C558-H578, C586-H606	MOTIFS
17	7487977CD1	240	S39 S50 S120 S137 T227 Y77	N203	HMG (high mobility group) box: V8-K76	HMMER_PFAM
					HMG1/2 proteins BL00353: R41-L87	BLIMPS_BLOCKS
					TRANSCRIPTION PROTEIN DN PD02448: N13-E51, A52-L99, P156-T169	BLIMPS_PRODUM
					PROTEIN DNA-BINDING NUCLEAR TRANSCRIPTION FACTOR REGULATION MOBILITY GROUP HIGH REPEAT PD000156: K9-Y77	BLAST_PRODUM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					HMG BOX DM00056 P47792 42-114: P4-K70 DM00056 P48430 28-100: D6-K70 DM00056 P53783 40-112: D6-K70 DM00056 P48433 35-107: D6-K70	BLAST_DOMO
18	1706514CD1	555	S42 S56 S134 S154 S318 S347 S403 S406 S459 S512 S518 T14 T36 T110 T189 T223 T322 T378 T434 Y196 Y537	N52 N73 N77	KRAB box: V4-E64	HMMER_PFAM
					Zinc finger, C2H2 type: H448-H470, H504-H526, Y224-H246, Y196-H218, Y308-H330, Y476-H498, Y420-H442, Y252-H274, Y364-H386, Y280-H302, Y392-H414, Y336-H358, Y140-H162	HMMER_PFAM
					Zinc finger, C2H2 type BL00028: C310-H326	BLIMPS_BLOCKS
					C2H2-type zinc finger signature PR00048: P307-S320, L323-G332	BLIMPS_PRINTS
					PROTEIN ZINC FINGER ZINC PD01066: F6-D44	BLIMPS_PRODOM
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING ZINC FINGER PATTERNALLY EXPRESSED ZN-FINGER PW1 PD017719: P251-H498, G276-F513, K300-K549, Y224-H470, R188-Q446, G136-Q390, E171-F401, E113-H330	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K248-C313, Y392-C453, K278-C341, K306-C369, K334-C397, K474-E533, H448-C509, K362-C425, K222-C285, K418-C481	BLAST_PRODROM
					HYPOTHETICAL ZINC FINGER PROTEIN B03B8.4 IN CHROMOSOME III ZINC FINGER DNA-BINDING METAL-BINDING NUCLEAR PD149420: K248-G416	BLAST_PRODROM
					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 Q05481 789-829: R468-Q508, H274-Q312, R244-Q283, R328-Q368 DM00002 P08042 314-358: C285-H330, C397-H442, C341-H386	BLAST_DOMO
					KRAB BOX DOMAIN DM00605 P52737 1-76: M1-K50	BLAST_DOMO
					Zn-finger, C2H2, type: C142-H162, C198-H218, C226-H246, C254-H274, C282-H302, C310-H330, C338-H358, C366-H386, C394-H414, C422-H442, C450-H470, C478-H498, C506-H526	MOTIFS
19	7488247CD1	184	S165 T143		PROTEIN NUCLEAR MSG-RELATED MRG1 MELANOCYTE-SPECIFIC ALTERNATIVE SPLICING PD015019: P117-C184	BLAST_PRODROM

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	1427269CD1	553	S21 S43 S50 S51 S95 S133 S187 S196 S214 S229 S268 S309 S411 S430 S433 T80 T148 T249 T282 T349 T417 T434	N15 N19 N227 N376	ELM2 domain: V57-A119	HMMER_PFAM
					Myb-like DNA-binding domain: N345-R390, P144-K189	HMMER_PFAM
					ATP/GTP-binding site motif A (P-loop): A158-S165	MOTIFS
21	103135CD1	371	S18 S28 S51 S58 S119 S242 S261 S324 T9 T154 T212 T317 T340 T365 Y260	N40 N225 N302	KRAB box: V8-D70	HMMER_PFAM
					Zinc finger, C2H2 type: Y260-H282, F204-H226, F288-H310, Y344-H366, Y316-H338, F232-H254, F176-H198	HMMER_PFAM
					C2H2-type zinc finger signature: PR00048: L303-G312, P259-K272	BLIMPS_PRINTS
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN TRANSCRIPTION REGULATION PD001562: V8- V68	BLAST_PRODROM
					PROTEIN ZINC FINGER METAL-BINDING DNA- BINDING PATERNALLY EXPRESSED PW1 PD017719: F176-F353	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					KRAB BOX DOMAIN DM00605 P17097 1-76: E6-E75 DM00605 I48689 11-85: Q5-P74 DM00605 P52738 3-77: Q5-E72 DM00605 P51523 5-79: Q5-P74	BLAST_DOMO
					Zn-finger, C2H2 type: C178-H198, C206-H226, C233-H254, C234-H254, C262-H282, C290-H310, C318-H338, C346-H366	MOTIFS
22	1907346CD1	837	S37 S77 S126 S149 S251 S293 S426 S538 S573 S622 S734 T28 T87 T116 T248 T264 T314 T349 T527 T566 T657 T685 T747	N232 N290 N318 N374 N682	KRAB box: L27-K89	HMMER_PFAM
					Zinc finger, C2H2 type: F388-H410, Y360-H382, Y416-H438, Y612-H634, F500-H522, Y472-H494, Y444-H466, Y668-H690, Y556-H578, Y276-H298, Y696-H718, Y584-H606, Y752-H774, Y304-H326, Y808-H830, Y780-H802, Y528-H550, Y724-H746, Y332-H354, Y640-H6	HMMER_PFAM
					C2H2-type zinc finger signature PR00048: P275-N288, L599-G608	BLIMPS_PRINTS
					PROTEIN ZINC FINGER PD01066: F29-G67 PD00066: H462-C474	BLIMPS_PRODROM
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING PATERNALLY EXPRESSED PW1 PD017719: G580-H830, G412-H662, P275-H522	BLAST_PRODROM

Table 3

SSEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: L27-K89 PD000072: K554-C617, K694-C757, K526-C589 MYELOBLAST KIAA0211 ZINC FINGER METAL-BINDING DNA-BINDING PD149061: K585-H770 KRAB BOX DOMAIN DM00605 P52738 3-77: Q24-K93 DM00605 Q05481 10-83: L27-K89 DM00605 Q03923 1-75: L27-P98 ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 Q05481 831-885: C645-Q700, C337-E392 Zn-finger, C2H2 type: C278-H298, C306-H326, C334-H354, C362-H382, C390-H410, C418-H438, C446-H466, C474-H494, C502-H522, C530-H550, C558-H578, C586-H606, C614-H634, C642-H662, C670-H690, C698-H718, C726-H746, C754-H774, C782-H802, C810-H830	BLAST_PRODOM
					KRAB BOX DOMAIN DM00605 P52738 3-77: Q24-K93 DM00605 Q05481 10-83: L27-K89 DM00605 Q03923 1-75: L27-P98 ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 Q05481 831-885: C645-Q700, C337-E392	BLAST_PRODOM
					Zn-finger, C2H2 type: C278-H298, C306-H326, C334-H354, C362-H382, C390-H410, C418-H438, C446-H466, C474-H494, C502-H522, C530-H550, C558-H578, C586-H606, C614-H634, C642-H662, C670-H690, C698-H718, C726-H746, C754-H774, C782-H802, C810-H830	MOTIFS
23	3041036CD1	549	S103 S173 S256 S287 T15 T24 T57 T157 T290 T452	N170 N260 N368 N379 N426 N510 N540	KRAB box: V14-K74	HMMER_PFAM
					Zinc finger, C2H2 type: F218-H240, Y442-H464, Y526-H548, Y246-H268, F414-H436, Y386-H408, Y302-H324, Y470-H492, Y498-H520, Y274-H296, Y358-H380, Y190-H212, Y330-H352	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN ZINC FINGER ZINC PD01066: F16-D54	BLIMPS_PRODUM
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING PATTERNALLY EXPRESSED PW1: PD017719: G270-H520, Y190-H436	BLAST_PRODUM
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN ZINC FINGER F18547_1 R28830_2 TRANSCRIPTION REGULATION PD009300: M343-Y442	
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K328-C391	
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: V14-L53	
					ZINC FINGER, C2H2 TYPE, DOMAIN DM000002 P08042 314-358: C503-H548 DM000002 Q05481 789-829: Q489-E530 DM000002 P52743 31-93: L485-H548	BLAST_DOMO
					KRAB BOX DOMAIN DM00605 I48689 11-85: E12-M61	BLAST_DOMO
					Zn-finger, C2H2 type: C192-H212, C220-H240, C248-H268, C276-H296, C304-H324, C332-H352, C360-H380, C388-H408, C416-H436, C444-H464, C472-H492, C500-H520, C528-H548	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
24	3856879CD1	555	S27 S61 S67 S129 S344 S372 S400 S456 S484 T18 T106 T119 T363 Y418	N154 N316	KRAB box: L17-D79	HMMER_PFAM
					Zinc finger, C2H2 type: H334-H356, Y362-H384, H278-H300, Y446-H468, Y390-H412, Y222-H244, F474-H496, Y306-H328, Y250-H272, Y194-H216, Y418-H440, Y502-H524, Y530-H552	HMMER_PFAM
					C2H2-type zinc finger signature PR00048: P221-S234, L293-G302	BLIMPS_PRINTS
					PROTEIN ZINC FINGER ZINC PD01066: F19-G57	BLIMPS_PRODUM
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING PATERNALLY EXPRESSED PW1 PD017719: G190-H440, G302-F539	BLAST_PRODUM
					HYPOTHETICAL ZINC FINGER PROTEIN B03B8.4 IN CHROMOSOME III DNA-BINDING METAL-BINDING NUCLEAR PD149420: E303-G470	BLAST_PRODUM
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: L17-D79	BLAST_PRODUM
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K472-C535, K360-C423, K332-C395, K220-C283, R192-C255, K388-C451, K444-C507	BLAST_PRODUM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					KRAB BOX DOMAIN DM00605 P52738 3-77: Q14-K85 DM00605 I48689 11-85: Q14-K81 DM00605 P52736 1-72: L17-K81 DM00605 P17097 1-76: L17-Q87	BLAST_DOMO
					Zn-finger, C2H2 type: C196-H216, C224-H244, C252-H272, C280-H300, C308-H328, C336-H356, C364-H384, C392-H412, C420-H440, C448-H468, C476-H496, C504-H524, C532-H552	MOTIFS
25	4178665CD1	601	S30 S31 S44 S67 S81 S87 S91 S98 S108 S155 S223 S502 T26 T51 T92 T207 T235 T263 T304 T416 T444 Y436	N65	Signal cleavage: M1-S22	SPSCAN
					Zinc finger, C2H2 type: Y380-H402, Y408-H430, C156-H178, F352-H374, Y436-H458, Y212-H234, H128-H150, Y240-H262, Y520-H542, Y464-H486, Y548-H570, Y268-H290, Y492-H514, Y324-H346, Y184-H206, Y296-H318, Y576-H598	HMMER_PFAM
					Transmembrane domain: M1-K25; N-terminus is non-cytosolic	TMAP
					C2H2-type zinc finger signature PR00048: P519-S532, L283-G292	BLIMPS_PRINTS
					PROTEIN ZINC FINGER, METAL-BINDING PD00066: H370-C382	BLIMPS_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING PATERNALLY EXPRESSED PW1 PD017719: G236-H486, G180-F417	BLAST_PRODROM
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR TRANSCRIPTION REGULATION REPEAT PD000072: K462-C525	BLAST_PRODROM
					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 Q05481 789-829: Q539-K579 DM00002 Q05481 831-885: C525-P575 DM00002 P08042 272-312: Q539-C578 DM00002 P08042 314-358: C385-H430	BLAST_DOMO
					Zn-finger, C2H2 type: C130-H150, C156-H178, C158-H178, C186-H206, C214-H234, C242-H262, C270-H290, C298-H318, C326-H346, C354-H374, C382-H402, C410-H430, C438-H458, C466-H486, C494-H514, C522-H542, C550-H570, C578-H598	MOTIFS
26	7493326CD1	743	S52 S58 S111 S161 S253 S337 S421 S618 S673 T9 T18 T68 T69 T264 T274 T371 T393 T412 T477 T505 T561 T644 T649 T659 T702 T715 Y271 Y720	N40 N105 N151 N198	KRAB box: L8-E70	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Zinc finger, C2H2 type: F383-H405, Y411-H433, Y636-H658, H215-H237, Y355-H377, Y299-H321, Y720-H742, F664-H686, Y523-H545, Y692-H714, Y551-H573, Y495-H517, H243-H265, Y439-H461, Y608-H630, Y467-H489, Y271-H293, H327-H349	HMME PFAM
					ZINC FINGER, C2H2-type BL00028: C245-H261	BLIMPS_BLOCKS
					C2H2-type zinc finger signature PR00048: P691-K704, L454-K463	BLIMPS-PRINTS
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING PATTERNALLY EXPRESSED PW1 PD017719: P494-H742, K347-L589, S200-F420	BLAST_PRODROM
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: L8-E65	BLAST_PRODROM
					KRAB BOX DOMAIN DM00605 I48689 11-85: Q5-C78 DM00605 P51523 5-79: Q5-S81 DM00605 P52736 1-72: L8-C78 DM00605 P52738 3-77: Q5-I73	BLAST_DOMO
					Zn-finger, C2H2 type: C217-H237, C245-H265, C273-H293, C301-H321, C329-H349, C357-H377, C385-H405, C413-H433, C441-H461, C469-H489, C497-H517, C525-H545, C553-H573, C610-H630, C638-H658, C694-H714, C722-H742	MOTIFS

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
27	1553836CD1	490	S12 S21 S77 S159 S265 S423 S426 S443 S462 T209 T397 Y36 Y73 Y326		Zinc finger, C2H2 type: L210-H233, Y363-H386, F120-H143, I329-C353, L174-C198, V92-C115	HMMER_PFAM
					Zn-finger, C2H2 type: C122-H143, C212-H233, C365-H386	MOTIFS
28	1908201CD1	665	S148 S161 S214 S241 S273 S317 S336 S340 S349 S350 S369 S391 S457 S545 S546 S578 S587 S597 S657 T38 T67 T86 T173 T184 T186 T206 T266 T343 T443 T472 T496 T528 T631 T639 Y263	N271 N355 N398 N402 N435 N500 N614 N643 N651	Zinc finger, C2H2 type: H618-H640, L486-H508, Y519-H542, C359-H382, L230-H253, F69-C92, R287-H310	HMMER_PFAM
					Zn-finger, C2H2 type: C232-H253, C289-H310, C359-H382, C361-H382, C488-H508, C620-H640	MOTIFS
29	2827615CD1	570	S16 S51 S57 S90 S91 S99 S118 S193 S198 S363 S529 T7 T95 T146 T170	N14 N160 N223 N336 N448	signal_cleavage: M1-P53	SPSCAN
					KRAB box: V6-K69	HMMER_PFAM
					Zinc finger, C2H2 type: Y521-H543, Y409-H431, F325-H347, Y353-H375, Y493-H515, Y241-H263, Y269-H291, Y465-H487, Y297-H319, L213-H235, Y381-H403, Y437-H459	HMMER_PFAM
					C2H2-type zinc finger domain PR00048: P436-S449, L228-G237	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING PATERNALLY EXPRESSED PW1 PD017719: G237-H487, G377-L562	BLAST_PRODUM
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: V6-K69	BLAST_PRODUM
					ZINC-FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT: PD000072: K435-C498	BLAST_PRODUM
					KRAB BOX DOMAIN DM00605 P52738 3-77: V6-R74 DM00605 I48208 18-93: V6-R74 DM00605 S42077 18-93: V6-R74 DM00605 I48689 11-85: K3-P78	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): A242-T249	MOTIFS
					Zn-finger, C2H2 type: C215-H235, C243-H263, C271-H291, C299-H319, C327-H347, C355-H375, C383-H403, C411-H431, C439-H459, C467-H487 C495-H515 C523-H543	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
30	4304550CD1	1712	S55 S102 S133 S176 S200 S232 S271 S286 S325 S443 S448 S504 S550 S561 S570 S607 S622 S629 S678 S789 S922 S933 S951 S971 S992 S1026 S1075 S1190 S1269 S1323 S1377 S1399 S1472 S1578 S1626 S1637 S1682 T115 T301 T432 T558 T576 T662 T1156 T1225 T1468 T1563 T1675 T1704 Y94 Y1072 Y1666	N114 N237 N778 N947 N1617	DEAD/DEAH box helicase: P764-S971	HMMER_PFAM
					HYPOTHETICAL HELICASE C28H8.3 IN CHROMOSOME III PROTEIN ATP-BINDING NUCLEAR PD135267: S1578-K1711, R1329-K1545	BLAST_PRODUM
					do SK12W; SK12; NUCLEOLAR; HELICASE DM01537 P47047 131-583: C1213-F1362, I748-L931 DM01537 A56003 60-514: D1223-V1337, P755-K935 DM01537 S56752 289-744: D1223-V1337, P755-K935, L1117-T1148 DM01537 P35207 309-803: P755-K935, K1254-V1337, F1114-K1157	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): A785-T792	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
31	7473738CD1	780	S80 S140 S387 S395 S403 S431 S486 S753 T43 T309 T438 T664 T677 T725 T775 Y256 Y533	N163 N347	DEAH-box subfamily ATP-dependent helicases proteins BL00690: G76-Q85, T104-E121, I170-T179	BLIMPS_BLOCKS
					DEAD and DEAH box families ATP-dependent helicases signatures: Y148-S197	PROFILES SCAN
					POLYPROTEIN PROTEIN HELICASE GENOME RNA CONTAINS: NUCLEAR ENVELOPE ATP-BINDING NONSTRUCTURAL PD000440: I316-S411, T49-P198, P336-I409	BLAST_PROD OM
					HELICASE RNA ATP-BINDING PROTEIN ATP-DEPENDENT NUCLEAR SPLICING mRNA PROCESSING PREmRNA PD001259: C406-D553	BLAST_PROD OM
					HELICASE PD091835: L72-A350	BLAST_PROD OM
					DEAH-BOX SUBFAMILY Y ATP-DEPENDENT HELICASES DM00649 P15938 341-962: T49-Y681 DM00649 P53131 84-705: P54-Y681 DM00649 P45018 50-661: T52-L293, D250-Y681 DM00649 G64165 74-685: T52-L293, D250-Y681	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): G76-T83	MOTIFS
					DEAH-box subfamily ATP-dependent helicases signature: S168-E177	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
32	4447743CD1	648	S14 S48 S160 S298 S326 S494 S619 T5 T54 T229 T241 T270 T397 T425 T466 T593 T643	N210 N350 N434 N518 N574 N602	Zinc finger, C2H2 type: Y256-H278, Y396-H418, Y452-H474, Y536-H558, Y564-H586, F592-H614, Y508-H530, Y312-H334, Y228-H250, Y284-H306, Y480-H502, F173-H194, F145-H167, Y620-H642, Y368-H390, Y200-H222, Y424-H446, Y340-H362	HMMER_PFAM
					KRAB box: V4-K66	HMMER_PFAM
					Zinc finger, C2H2 type BL00028: C398-414	BLIMPS_BLOCKS
					Zinc ribbon domain, C2H2 type BL00466: C342-T378	BLIMPS_BLOCKS
					C2H2-type zinc finger signature PR00048: P255-S268, L579-G588	BLIMPS_PRINTS
					PROTEIN ZINC FINGER ZINC PD01066: F6-G44	BLIMPS_PRODOM
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN ZINC FINGER NUCLEAR TRANSCRIPTION REGULATION REPEAT PD008015: H68-G169	BLAST_PRODOM
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING ZINC FINGER PATTERNALLY EXPRESSED ZN-FINGER PW1 PD017719: G224-H474	BLAST_PRODOM
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: V4-P62	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K198-C261, K422-C485, K366-C429, K534-C597	BLAST_PRODOM
					KRAB BOX DOMAIN DM00605 Q05481 10-83: M1-P74 DM00605 Q03923 1-75: V4-P74 DM00605 P28160 1-69: D8-P74 DM00605 S22564 1-63: F13-P74	BLAST_DOMO
					Zn-finger, C2H2 type: C147-H167, C202-H222, C230-H250, C258-H278, C286-H306, C314-H334, C342-H362, C370-H390, C398-H418, C426-H446, C454-H474, C482-H502, C510-H530, C538-H558, C566-H586, C594-H614, C622-H642	MOTIFS
33	7497554CD1	602	S2 S7 S12 S26 S76 S108 S109 S179 S305 S460 S525 S560 T74 T82 T91 T151 T222 T310 T330	N308 N395	T-box: Q114-S305	HMMER_PFAM
					T-box domain proteins BL01283: L122-D169, W182-N223, L234-I247, F272-D304	BLIMPS_BLOCKS
					T-box domain signature PR00937: I233-I247, T275-I288, N296-D304, G130-D154, Y197-M210, V214-N223	BLIMPS_PRINTS
					Brachyury protein family signature PR00938: Y177-Y197, H238-L255	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					TBX15 PROTEIN T-BOX 15 TRANSCRIPTION REGULATION DNA-BINDING NUCLEAR PD040336: G306-V602 PD102400: M1-E112	BLAST_PRODROM
					PROTEIN DNA-BINDING NUCLEAR TRANSCRIPTION T-BOX REGULATION DEVELOPMENTAL BRACHYURY ACTIVATOR PD001585: L153-S305 PD001407: M107-G152	BLAST_PRODROM
					T-BOX DM01478 S46458 62-380: E111-S362 DM01478 A40213 294-606: M110-R309 DM01478 A49125 1-329: S108-S305 DM01478 P24781 6-337: E111-E317	BLAST_DOMO
					T-box domain signature 1: L122-R141	MOTIFS
					T-box domain signature 2: I198-F216	MOTIFS
34	7475843CD1	388	S16 S87 S168 T22 T220 Y151	N102 N235 N295	KRAB box: M1-K34	HMMER_PPFAM
					Zinc finger, C2H2 type: Y281-H303, Y309-H331, Y337-H359, Y197-H219, H253-H275, Y225-H247, Y169-H191, F113-H135, Y365-H387	HMMER_PPFAM
					Zinc finger, C2H2 type BL00028:C283-H299	BLIMPS_BLOCKS
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN ZINC FINGER NUCLEAR TRANSCRIPTION REGULATION REPEAT PD008015: K35-G137	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING ZINC FINGER PATTERNALLY EXPRESSED ZN-FINGER PW1 PD017719: N116-I386	BLAST_PRODOM
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K279-C342, K195-C258	BLAST_PRODOM
					ZINC FINGER PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING PUTATIVE REX2 TRANSCRIPTION REGULATION	BLAST_PRODOM
					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 Q0548 831-885: C286-E341, C202-E257 DM00002 Q0548 789-829: K300-E341, K244-E285 DM00002 P52743 31-93: L296-H359 DM00002 Q0548 887-927: C258-H299	BLAST_DOMO
					Zn-finger, C2H2 type: C171-H191, C199-H219, C227-H247, C255-H275, C283-H303, C311-H331, C339-H359	MOTIFS
35	6319550CD1	480	S141 S160 S229 S243 S244 S288 S290 S315 S337 S346 T98 T427		BTB/POZ domain: S22-Y135	HMMER_PFAM
					Zinc finger, C2H2 type: Y357-H379, Y413-H435, Y385-H407, Y329-H351, Y441-H464	HMMER_PFAM
					C2H2-type zinc finger signature PR00048: P384-V397, L428-G437	BLIMPS_PRINTS
					BTB domain PF00651: A51-F63	BLIMPS_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					B-CELL DNA-BINDING ZINC FINGER METAL-BINDING LYMPHOMA PROTEIN BCL6 NUCLEAR TRANSCRIPTION REGULATION PD025107: E148-C313 PD019050: K439-I476	BLAST_PRODOM
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K327-C390	BLAST_PRODOM
					POZ DOMAIN DM00509 P41182 7-213: E16-E148 DM00509 S59069 1-171: E16-P137 DM00509 P10074 1-153: F17-F130	BLAST_DOMO
					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 P41182 649-679: V432-K463	BLAST_DOMO
					Zn-finger, C2H2 type: C331-H351, C359-H379, C387-H407, C415-H435, C443-H464	MOTIFS
36	7510064CD1	790	S46 S54 S69 S164 S172 S196 S209 S309 S317 S386 S404 S414 S440 S509 S547 S553 S581 S641 S662 S663 S708 T204 T217 T229 T296 T518 T535 T658 Y308	N156 N234 N589	Signal Peptide: M1-T24	HMMER
					helix-loop-helix domain: R64-W119	HMMER_SMART
					PAS domain: P128-T194, P271-S339	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Nuclear translocator signature PR00785: A93-L113, V143-G162, D175-Q193, W230-S249, A284-R301	BLIMPS_PRINTS
					PROTEIN NUCLEAR REPEAT TRANSCRIPTION REGULATION DNABINDING RECEPTOR HYDROCARBON ARYL BIOLOGICAL PD008299: E205-V388	BLAST_PRODOM
					PROTEIN TRANSCRIPTION REGULATION DNA-BINDING REPEAT NUCLEAR RECEPTOR HYDROCARBON ARYL DEVELOPMENTAL PD002640: K60-G117	BLAST_PRODOM
					AH; MINDED; SINGLE; HYPOXIA; DM03763 I38972 18-524: R59-E401 DM03763 P05709 1-564: K60-T190, E205-V498, K60-T190 DM03763 P35869 28-512: K62-T204, E205-R371 DM03763 P30561 27-503: K62-R371, P373-G392	BLAST_DOMO

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
37/7490148CB1/ 1154	1-323, 71-368, 204-629, 211-667, 224-629, 226-599, 235-1145, 235-1148, 286-599, 289-629, 705-1143, 705-1154, 734-1143
38/7490301CB1/ 754	1-299, 1-754, 154-705, 155-705
39/2383223CB1/ 2483	1-203, 1-559, 1-563, 1-623, 1-652, 3-476, 4-125, 4-238, 4-401, 4-474, 4-564, 4-659, 16-601, 45-666, 49-465, 74-327, 74-682, 122-697, 144-272, 332-945, 332-946, 339-946, 393-937, 440-1010, 498-1048, 526-567, 526-590, 526-604, 526-617, 526-627, 526-628, 526-638, 526-649, 526-655, 526-675, 526-726, 526-728, 526-735, 526-779, 526-813, 526-817, 526-819, 526-899, 526-905, 529-796, 535-613, 535-649, 535-817, 535-901, 550-621, 550-652, 550-901, 550-1037, 551-899, 578-816, 588-628, 588-647, 588-648, 588-719, 588-726, 588-819, 588-869, 588-897, 588-900, 589-638, 589-728, 589-880, 600-650, 600-901, 600-904, 600-931, 600-1131, 601-899, 604-730, 604-817, 604-1067, 607-899, 613-997, 613-1037, 613-1065, 614-795, 614-962, 615-940, 619-726, 619-901, 619-1132, 662-901, 663-1067, 672-722, 672-815, 672-873, 672-880, 672-899, 673-817, 682-711, 682-719, 682-728, 684-726, 684-745, 684-899, 690-905, 690-1196, 691-901, 705-728, 715-1069, 715-1146, 749-1196, 757-817, 757-836, 757-962, 757-1068, 757-1069, 757-1145, 759-797, 768-817, 768-1067, 768-1141, 768-1145, 769-899, 769-1067, 769-1106, 769-1156, 772-894, 772-901, 772-1146, 781-902, 781-1153, 781-1300, 783-899, 784-901, 787-901, 787-1067, 787-1118, 789-819, 791-1146, 796-1153, 796-1289, 803-901, 824-899, 833-899, 839-870, 839-893, 840-899, 840-1048, 847-879, 847-1321, 852-1235, 853-904, 856-899, 862-1239, 865-1239, 865-1289, 866-1041, 866-1216, 917-1215, 924-1076, 933-1131, 933-1132, 933-1145, 943-1062, 943-1141, 943-1230, 943-1235, 943-1391, 951-994, 951-1047, 951-1145, 951-1235, 964-1321, 964-1403, 991-1067, 993-1391, 1003-1153, 1003-1235, 1003-1496, 1007-1038, 1007-1391, 1008-1069, 1008-1314, 1008-1321, 1011-1049, 1015-1069, 1020-1457, 1024-1146, 1027-1230, 1027-1314, 1030-1133, 1030-1153, 1030-1404, 1030-1409, 1030-1500, 1030-1552, 1034-1145, 1036-1230, 1048-1393, 1048-1541, 1085-1496, 1086-1699, 1093-1146, 1093-1235, 1093-1314, 1099-1141, 1102-1347, 1104-1314, 1105-1144, 1105-1146, 1105-1573, 1107-1230, 1114-1301, 1114-1405, 1114-1492, 1114-1636, 1119-1456, 1140-1700, 1169-1209, 1169-1235, 1171-1216, 1189-1237, 1191-1412, 1192-1229, 1192-1230, 1192-1235, 1192-1237, 1192-1239, 1192-1313, 1192-1321, 1192-1377, 1192-1384, 1192-1393, 1192-1571, 1198-1235, 1198-1301, 1198-1321, 1198-1398, 1198-1571, 1198-1720, 1199-1235, 1204-1314, 1216-1299, 1216-1566, 1222-1289, 1223-1700, 1253-1409, 1255-1299, 1255-1314, 1255-1467, 1255-1468, 1255-1483,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1255-1553, 1255-1569, 1255-1575, 1272-1321, 1272-1643, 1276-1313, 1276-1314, 1276-1741, 1312-1909, 1351-1378, 1351-1384, 1351-1403, 1351-1404, 1351-1405, 1351-1551, 1351-1552, 1351-1575, 1351-1637, 1351-1649, 1351-1653, 1351-1719, 1351-1743, 1351-1793, 1356-1385, 1356-1393, 1356-1405, 1356-1406, 1356-1655, 1356-1670, 1356-1687, 1356-1753, 1356-1793, 1360-1403, 1360-1406, 1369-1774, 1369-1805, 1369-1821, 1370-1551, 1370-1720, 1387-1719, 1391-1593, 1441-1573, 1463-1824, 1507-1566, 1507-1643, 1507-1824, 1507-1827, 1512-1542, 1512-1571, 1512-1573, 1512-1657, 1512-1739, 1512-1782, 1512-1825, 1512-1828, 1513-1566, 1513-1687, 1515-1566, 1519-1571, 1519-1823, 1519-1887, 1521-1733, 1522-1571, 1524-1825, 1525-1827, 1528-1575, 1528-1625, 1528-1637, 1528-1729, 1534-1575, 1534-1663, 1534-1827, 1534-1887, 1537-1887, 1538-1713, 1538-1739, 1547-1827, 1552-1887, 1559-1729, 1596-1741, 1596-1793, 1596-1827, 1596-1866, 1603-1635, 1603-1827, 1608-1733, 1609-1887, 1612-1657, 1618-1678, 1618-1721, 1618-1730, 1618-1767, 1618-1820, 1618-1827, 1621-1827, 1631-1820, 1671-1823, 1671-1825, 1672-1739, 1687-1743, 1693-1733, 1695-1743, 1696-1733, 1696-1734, 1696-1739, 1696-1743, 1696-1828, 1702-1818, 1702-1825, 1708-1733, 1770-1827, 1779-2060, 1896-2160, 1936-2064, 2039-2306, 2039-2458, 2039-2482, 2039-2483
40/3495982CB1/ 2535	1-313, 1-367, 96-364, 96-366, 96-750, 165-610, 298-355, 354-881, 354-898, 354-1029, 494-747, 513-1156, 521-1435, 529-1254, 537-1220, 538-1199, 539-1217, 551-959, 560-979, 702-828, 721-1258, 735-1214, 745-1359, 750-1270, 761-795, 761-814, 761-818, 761-825, 761-828, 761-904, 761-995, 761-1072, 761-1080, 767-904, 768-827, 768-828, 768-914, 768-965, 769-818, 769-828, 771-965, 774-1080, 775-818, 780-824, 780-827, 783-828, 787-1080, 791-965, 801-1258, 804-1107, 809-1080, 839-1358, 844-965, 845-902, 845-912, 845-965, 845-1079, 845-1164, 854-1372, 864-904, 864-965, 864-1080, 868-904, 871-1269, 888-1164, 893-1399, 904-1523, 934-1129, 936-968, 936-993, 936-995, 936-1073, 936-1080, 936-1133, 936-1134, 939-1047, 948-1164, 959-1522, 960-1233, 965-1613, 995-1516, 1002-1350, 1012-1613, 1013-1070, 1013-1080, 1013-1156, 1020-1072, 1020-1073, 1020-1079, 1020-1164, 1022-1073, 1029-1210, 1030-1072, 1030-1080, 1032-1079, 1032-1133, 1032-1164, 1039-1133, 1040-1133, 1064-1133, 1087-1588, 1094-1588, 1105-1164, 1114-1161, 1114-1164, 1116-1156, 1116-1164, 1117-1156, 1119-1644, 1131-1645, 1170-1735, 1189-1766, 1196-1847, 1219-1566, 1219-1570, 1220-2089,

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
	1221-2103, 1222-2098, 1226-1804, 1227-1599, 1228-1776, 1228-1859, 1230-1820, 1230-1890, 1231-1850, 1231-1851, 1233-1624, 1233-1668, 1233-1838, 1236-1854, 1243-1833, 1245-1849, 1250-1942, 1266-1882, 1267-2077, 1271-2078, 1287-1617, 1287-2088, 1290-1965, 1305-1842, 1307-1918, 1309-1636, 1315-1672, 1322-2079, 1329-1893, 1343-1704, 1347-1971, 1355-1807, 1363-2089, 1380-2009, 1386-1933, 1399-1833, 1404-1656, 1408-2068, 1449-2054, 1465-2103, 1492-2114, 1496-2074, 1506-1656, 1519-1731, 1525-1579, 1525-1589, 1527-2090, 1580-2060, 1586-2025, 1633-1827, 1633-2100, 1633-2130, 1633-2135, 1645-1916, 1645-1932, 1645-2027, 1645-2085, 1645-2124, 1645-2166, 1645-2207, 1645-2208, 1645-2235, 1645-2239, 1645-2268, 1645-2288, 1648-2166, 1655-1894, 1698-1952, 1699-1973, 1702-2220, 1726-2138, 1750-1900, 1767-2395, 1835-2518, 1864-2501, 1870-2302, 1871-2080, 1871-2120, 1871-2416, 1903-2508, 1929-2506, 1930-2515, 1948-2317, 1965-2153, 1976-2493, 1982-2419, 2048-2484, 2059-2361, 2059-2535, 2104-2521, 2106-2535, 2111-2535, 2112-2371, 2118-2341, 2147-2518, 2177-2518, 2191-2533, 2196-2530, 2196-2535, 2249-2361, 2336-2518
41/7477891CBI/ 2817	1-258, 1-282, 1-450, 1-489, 1-507, 1-511, 1-596, 1-738, 2-572, 3-513, 60-322, 60-663, 60-719, 60-733, 60-738, 61-589, 63-735, 78-290, 146-414, 163-738, 164-675, 244-691, 284-724, 311-738, 344-738, 361-738, 486-944, 486-949, 486-994, 486-999, 486-1013, 486-1112, 486-1118, 486-1133, 486-1167, 486-1178, 486-1372, 502-1354, 546-827, 546-1165, 548-1092, 566-1233, 614-1516, 679-1354, 685-1354, 757-1332, 787-1518, 848-1518, 859-890, 859-925, 872-1518, 1018-1518, 1090-1518, 1106-1484, 1123-1518, 1126-1516, 1130-1518, 1134-1392, 1145-1515, 1203-1518, 1231-1507, 1241-1518, 1243-1518, 1247-1518, 1327-1518, 1342-1518, 1357-1518, 1360-1518, 1439-1654, 1441-1518, 1442-1518, 1454-1510, 1454-1511, 1454-1516, 1454-1528, 1454-1531, 1454-1542, 1454-1552, 1454-1558, 1454-1568, 1454-1586, 1454-1602, 1454-1603, 1454-1607, 1454-1623, 1454-1631, 1454-1635, 1454-1639, 1454-1655, 1454-1686, 1459-1686, 1460-1516, 1461-1686, 1465-1518, 1470-1565, 1470-1566, 1470-1589, 1470-1593, 1470-1612, 1470-1615, 1470-1632, 1470-1648, 1470-1655, 1470-1658, 1470-1668, 1470-1678, 1470-1679, 1470-1696, 1470-1699, 1470-1710, 1470-1720, 1470-1726, 1470-1736, 1470-1754, 1470-1770, 1470-1771, 1470-1775, 1470-1791, 1470-1799, 1470-1803, 1470-1807, 1470-1822, 1470-1823, 1470-1855, 1470-1857, 1470-1874, 1470-1885, 1470-1890, 1470-1892, 1470-1918, 1470-1921, 1470-1939, 1470-1948, 1470-1959, 1470-1961, 1470-1962, 1470-1963, 1470-1969, 1470-1970, 1470-1978, 1470-1983, 1470-1987, 1470-1989, 1470-1994, 1470-2004, 1470-2008, 1470-2017, 1470-2018, 1470-2033, 1470-2035, 1470-2039, 1470-2056, 1470-2072, 1470-2082, 1470-2085, 1470-2086, 1470-2097, 1470-2104, 1470-2110, 1470-2115, 1470-2117, 1470-2118, 1470-2128, 1470-2133, 1470-2183, 1470-2484, 1471-1516, 1472-1516, 1472-2128, 1475-1686, 1478-1686, 1481-1686, 1494-2135, 1496-2348, 1498-2177, 1500-1686, 1518-1686, 1524-1686, 1525-1686, 1529-1686, 1535-1686, 1536-2232, 1539-1686, 1540-1686, 1551-2532, 1553-1686, 1561-1686, 1568-1686, 1573-1686, 1574-1686, 1576-1686, 1579-1686, 1580-2293, 1602-1686, 1607-1684, 1608-2303, 1611-2240, 1614-2300, 1619-1686, 1620-2077, 1622-1684, 1627-1683, 1627-1684,

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
	1633-1686, 1639-1684, 1639-1733, 1639-1734, 1639-1757, 1639-1761, 1639-1780, 1639-1783, 1639-1800, 1639-1816, 1639-1823, 1639-1826, 1639-1836, 1639-1846, 1639-1847, 1639-1854, 1640-1854, 1643-1854, 1649-2212, 1650-2294, 1654-1854, 1661-2300, 1668-1854, 1681-2396, 1683-2222, 1686-2259, 1692-2306, 1697-2357, 1703-2353, 1704-1854, 1707-2243, 1708-2066, 1719-1854, 1721-2693, 1729-2602, 1736-2318, 1741-2291, 1742-1854, 1742-2434, 1744-2299, 1747-2250, 1748-1854, 1756-2468, 1770-2236, 1775-1852, 1776-1854, 1779-1854, 1787-2303, 1788-1854, 1790-1852, 1795-1851, 1795-1852, 1801-2131, 1807-1847, 1807-1852, 1817-1854, 1820-2596, 1841-2371, 1851-2411, 1854-2076, 1854-2078, 1854-2109, 1854-2122, 1854-2283, 1854-2394, 1860-2453, 1865-2095, 1869-2508, 1887-2135, 1889-2482, 1906-2604, 1913-2184, 1938-2482, 1954-2245, 1961-2434, 1968-2469, 1972-2003, 1972-2036, 1975-2233, 1977-2419, 1978-2817, 1980-2261, 1981-2276, 2001-2240, 2006-2255, 2008-2257, 2008-2263, 2008-2275, 2008-2295, 2008-2298, 2008-2487, 2008-2507, 2010-2308, 2013-2563, 2023-2175, 2023-2369, 2023-2413, 2023-2530, 2035-2503
42/72688352CB1/ 2295	1-543, 1-2277, 6-606, 10-555, 10-702, 14-698, 181-781, 203-781, 333-822, 633-1261, 679-1261, 706-1385, 707-967, 755-1267, 807-1017, 807-1238, 807-1322, 869-1110, 869-1559, 933-1586, 960-1439, 968-1554, 980-1181, 980-1302, 988-1215, 1008-1204, 1008-1215, 1008-1265, 1008-1288, 1008-1294, 1008-1302, 1008-1349, 1008-1456, 1008-1721, 1011-1294, 1012-1350, 1012-1352, 1012-1363, 1012-1887, 1013-1215, 1013-1294, 1022-1067, 1022-1095, 1022-1186, 1022-1187, 1022-1215, 1022-1224, 1022-1292, 1022-1301, 1022-1308, 1022-1352, 1022-1456, 1022-1630, 1022-1634, 1029-1456, 1029-1557, 1039-1308, 1039-1352, 1047-1239, 1072-1294, 1072-1456, 1073-1294, 1075-1783, 1078-1456, 1084-1363, 1084-1613, 1084-1630, 1084-1644, 1093-1263, 1093-1352, 1096-1345, 1096-1349, 1096-1447, 1096-1663, 1096-1971, 1101-1476, 1102-1447, 1107-1195, 1116-1456, 1117-1195, 1127-1714, 1129-1728, 1138-1294, 1138-1744, 1146-1644, 1147-1644, 1152-1697, 1156-1294, 1168-1193, 1168-1204, 1168-1344, 1168-1352, 1168-1383, 1168-1447, 1168-1467, 1170-1467, 1171-1363, 1173-1347, 1173-1438, 1173-2034, 1174-1204, 1176-1294, 1176-1456, 1176-1517, 1176-1624, 1176-1889, 1189-1303, 1230-1288, 1230-1349, 1230-1437, 1230-1447, 1230-1456, 1230-1728, 1230-1858, 1231-1781, 1239-1728, 1243-1363, 1243-1467, 1243-1569, 1243-1636, 1244-1467, 1248-1302, 1250-1302, 1250-1644, 1251-1467, 1251-1546, 1257-1431, 1257-1436, 1258-1546, 1262-1730, 1270-1302, 1271-1302, 1274-1362, 1274-1467, 1274-1490, 1274-1513, 1274-1519, 1274-1546, 1274-1549, 1274-1560, 1274-1604, 1274-1708, 1274-1882, 1275-1363, 1280-1560, 1280-1657, 1280-1887, 1292-1708, 1301-1491, 1315-1467, 1316-1638, 1318-1467, 1324-1546, 1324-1858, 1325-1546, 1327-1728, 1327-1971, 1330-1848, 1334-1447, 1354-1708, 1368-1638, 1369-1447, 1369-1638, 1369-1728, 1371-1882, 1379-1882, 1381-1882, 1398-1515, 1398-1517, 1398-1624, 1398-1889, 1399-1889, 1404-1882, 1407-1624, 1407-1889, 1408-1546, 1411-1447, 1411-1456, 1412-1624, 1413-1467, 1415-2205, 1417-1445, 1417-1456, 1417-1630, 1419-1456, 1424-1708, 1425-1598, 1425-1604, 1425-1882, 1426-1630, 1428-1546, 1428-1624, 1482-1881, 1483-1882, 1496-1708, 1502-1549, 1502-1552, 1502-1560, 1502-1629, 1502-1714, 1502-1721, 1503-1546, 1503-1550, 1503-1887, 1506-1629, 1514-1882, 1522-1633, 1522-1887, 1523-1887, 1529-2249,

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
	1532-1708, 1532-1882, 1538-1887, 1540-1887, 1558-1797, 1558-1887, 1567-1624, 1567-1638, 1567-1644, 1567-1728, 1576-1797, 1578-1992, 1580-2240, 1584-1638, 1587-1637, 1587-1797, 1587-1887, 1591-1887, 1600-1887, 1605-2289, 1606-1638, 1607-2022, 1618-2238, 1636-2274, 1637-2291, 1639-1894, 1658-2207, 1662-1888, 1666-2277, 1669-1708, 1669-1882, 1670-1720, 1670-1728, 1670-1797, 1670-1882, 1670-1889, 1671-1718, 1671-1721, 1671-1764, 1671-1876, 1671-1887, 1671-1966, 1671-1971, 1672-1708, 1672-1721, 1672-1887, 1672-1960, 1672-1971, 1674-2273, 1676-1709, 1678-1851, 1678-1856, 1678-1971, 1680-1721, 1680-1887, 1680-1925, 1680-1966, 1684-2167, 1689-1728, 1690-1966, 1694-1971, 1701-1965, 1701-1966, 1704-1936, 1714-2278, 1715-1977, 1740-2293, 1743-1887, 1744-1876, 1744-1887, 1744-1971, 1745-2291, 1745-2295, 1747-1971, 1755-1881, 1755-1971, 1758-1971, 1777-2209, 1778-1960, 1778-1971, 1785-1971, 1794-2295, 1805-2280, 1810-1966, 1819-1971, 1820-1971, 1822-2280, 1827-1971, 1828-1966, 1831-1882, 1831-1971, 1833-1887, 1833-1971, 1839-1877, 1839-1881, 1839-1882, 1839-1889, 1842-1889, 1845-1966, 1848-2114, 1848-2256, 1848-2282, 1863-2295, 1884-2278, 1917-1971, 1957-2277, 2004-2278, 2018-2292, 2088-2278, 2095-2232
43/7490652CB1/ 958	1-280, 104-847, 797-958
44/7489744CB1/ 1978	1-667, 516-661, 516-663, 516-673, 516-677, 516-707, 516-740, 516-771, 516-798, 516-933, 517-669, 517-810, 517-827, 517-929, 517-931, 517-1017, 518-812, 519-670, 519-671, 519-675, 519-677, 520-722, 520-752, 521-669, 521-671, 521-672, 521-673, 521-674, 521-675, 521-676, 521-680, 521-691, 521-714, 521-723, 521-725, 521-731, 521-792, 521-806, 521-812, 521-958, 521-977, 521-997, 521-1170, 522-672, 522-673, 522-676, 522-678, 522-679, 522-681, 522-690, 522-732, 522-740, 522-773, 522-785, 522-789, 522-837, 522-881, 522-897, 522-984, 522-1124, 523-660, 523-668, 523-671, 523-673, 523-674, 523-675, 523-676, 523-680, 523-689, 523-762, 523-780, 523-781, 523-902, 523-918, 523-958, 524-665, 524-673, 524-674, 524-675, 524-677, 524-678, 524-682, 524-683, 524-762, 524-774, 524-806, 524-940, 524-1087, 525-669, 525-673, 525-674, 525-678, 525-686, 525-708, 525-729, 525-767, 525-792, 525-797, 526-671, 526-672, 526-673, 526-675, 526-676, 526-678, 526-680, 527-671, 527-674, 527-676, 527-678, 527-680, 527-681, 527-683, 527-766, 527-776, 527-817, 528-674, 528-675, 528-678, 528-680, 528-692, 528-763, 528-881, 529-676, 529-680, 529-688, 529-729,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	529-857, 529-863, 529-935, 529-1035, 529-1126, 530-675, 530-678, 530-679, 530-686, 530-927, 530-1087, 531-675, 531-680, 531-690, 531-1164, 532-671, 532-673, 532-681, 532-758, 532-763, 532-926, 532-946, 532-961, 533-678, 534-675, 534-676, 534-677, 534-678, 534-679, 534-710, 535-682, 535-763, 536-729, 536-1978, 537-672, 537-680, 538-674, 538-694, 539-679, 539-814, 539-1067, 541-674, 541-676, 541-678, 541-683, 541-740, 541-745, 541-1009, 541-1111, 542-675, 542-678, 542-679, 542-683, 542-776, 543-674, 543-675, 543-679, 543-681, 543-945, 543-1091, 544-736, 544-816, 545-675, 545-676, 545-677, 545-686, 545-962, 546-677, 546-678, 546-679, 547-676, 547-678, 548-680, 548-956, 548-967, 548-1151, 550-679, 550-940, 550-967, 551-676, 551-738, 553-680, 556-721, 560-690, 560-759, 560-809, 560-886, 561-694, 561-1000, 583-1215, 612-968, 627-1384, 649-816, 651-1304, 659-963, 671-1124, 681-993, 695-1215, 697-1025, 704-1402, 708-1409, 709-926, 714-875, 716-910, 718-840, 720-1254, 721-1507, 722-952, 724-1275, 727-983, 728-1252, 728-1589, 732-927, 732-1001, 734-1324, 776-827, 777-859, 786-989, 883-939, 967-1381, 1030-1327, 1036-1188
45/3363382CB1/ 2859	1-397, 1-418, 87-916, 115-479, 167-533, 230-459, 230-454, 241-933, 368-551, 544-942, 544-1007, 544-1012, 544-1014, 544-1016, 544-1022, 544-1026, 545-997, 545-1007, 545-1020, 550-1127, 550-1211, 553-865, 553-932, 553-1002, 553-1042, 553-1113, 553-1146, 554-1112, 555-971, 562-695, 562-1010, 563-958, 567-1022, 592-1017, 629-1074, 710-845, 744-845, 783-1196, 816-988, 900-996, 900-1192, 907-1511, 1007-1652, 1049-1304, 1125-1597, 1180-1794, 1188-1452, 1206-1432, 1315-1555, 1316-1954, 1316-1977, 1398-1845, 1409-1852, 1484-2130, 1513-1801, 1513-1825, 1552-2040, 1565-1954, 1566-1817, 1566-2022, 1586-1740, 1599-1953, 1614-2016, 1629-1886, 1659-2287, 1664-2089, 1671-1803, 1696-1972, 1770-2008, 1776-2154, 1789-2335, 1848-2531, 1878-2102, 1929-2595, 1940-2524, 1952-2533, 1952-2536, 1952-2537, 1965-2533, 1980-2267, 2022-2620, 2024-2523, 2032-2403, 2047-2539, 2052-2538, 2057-2536, 2064-2538, 2077-2534, 2077-2538, 2098-2538, 2102-2608, 2122-2536, 2128-2525, 2142-2432, 2199-2607, 2234-2504, 2354-2528, 2482-2779, 2583-2799, 2592-2859
46/7491148CB1/ 1772	1-271, 1-1323, 270-547, 271-668, 343-980, 388-542, 393-542, 876-1492, 992-1461, 1007-1644, 1018-1502, 1064-1772, 1078-1623, 1079-1255, 1080-1267, 1080-1272, 1093-1696, 1172-1644, 1194-1653
47/8126343CB1/ 3112	1-1127, 82-571, 93-614, 104-427, 104-884, 105-966, 110-947, 119-584, 131-807, 131-817, 150-1005, 513-853, 570-1359, 783-1432, 838-1094, 854-1096, 854-1200, 854-1262, 883-1114, 890-1369, 915-1068, 938-1101, 1000-1249, 1139-1359, 1256-1695, 1364-1649, 1364-1771, 1364-1788, 1364-1840, 1364-1881, 1364-1898, 1364-1938, 1364-1969, 1364-1973, 1365-1472, 1425-2064, 1452-2026, 1468-2145, 1560-1763, 1570-2131, 1576-2107, 1576-2204, 1595-1778, 1641-2199, 1647-2230, 1740-2427, 1796-2506, 1815-2204, 1833-1951, 1834-2049, 1976-2538, 2048-2537, 2072-2287, 2072-2612, 2084-2635, 2099-2435, 2105-2605, 2122-2683, 2132-2347, 2132-2712, 2135-2465, 2170-2618, 2193-2494, 2193-2783, 2220-2898, 2233-2880, 2234-2451, 2254-2505, 2279-2884, 2340-2798, 2361-2645, 2371-2996, 2388-2694, 2406-2674, 2412-3081, 2434-2919, 2518-2998, 2522-2989, 2525-3112, 2538-3084, 2586-2783, 2586-2787, 2586-3112, 2903-2946, 3066-3097, 3066-3105, 3066-3109

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
48/7044055CB1/ 1939	1-427, 1-475, 1-543, 1-1931, 4-436, 6-735, 11-640, 11-650, 11-726, 12-509, 12-538, 33-485, 35-515, 40-640, 46-598, 76-721, 91-704, 164-448, 172-754, 208-762, 303-516, 331-885, 339-1039, 369-636, 384-973, 388-1049, 401-1182, 446-1153, 544-1122, 618-1370, 765-916, 769-1180, 802-1078, 842-1631, 852-1175, 852-1512, 900-1344, 1129-1628, 1154-1673, 1226-1851, 1250-1913, 1296-1899, 1307-1617, 1334-1939, 1342-1684, 1372-1928, 1382-1654, 1382-1917, 1402-1671, 1423-1690, 1447-1614, 1448-1829, 1448-1936, 1461-1853, 1482-1754, 1489-1925, 1539-1933, 1603-1877, 1604-1934
49/7493424CB1 5902	1-508, 1-633, 25-433, 51-788, 468-541, 586-1112, 598-859, 702-1112, 734-1021, 767-1404, 767-1461, 1102-1302, 1102-1502, 1126-1701, 1349-2038, 1388-1684, 1446-2299, 1542-2052, 1776-2311, 1780-2345, 1796-2400, 1889-2287, 1971-2239, 1971-2481, 1976-2251, 2054-2562, 2061-2790, 2094-2521, 2231-2836, 2231-2900, 2281-2559, 2282-2835, 2451-2963, 2471-2771, 2472-2827, 2489-3045, 2662-3072, 2880-3153, 2885-3176, 2887-3181, 2894-3017, 2935-3679, 3013-3137, 3035-3445, 3078-3563, 3155-3548, 3226-3585, 3237-3546, 3246-3491, 3255-3552, 3255-3556, 3255-3744, 3286-3701, 3312-3646, 3346-3731, 3346-3939, 3360-3605, 3406-3882, 3412-3546, 3450-3700, 3468-4051, 3514-3605, 3552-4363, 3720-4414, 3740-4337, 3745-4311, 3745-4391, 3757-4344, 3801-4387, 3974-4218, 3974-4242, 3981-4171, 4066-4328, 4067-4664, 4096-4392, 4106-4645, 4163-4460, 4170-4221, 4170-4473, 4181-4468, 4394-4679, 4406-4671, 4445-4691, 4445-4908, 4447-4700, 4458-4732, 4491-4739, 4549-4816, 4554-5360, 4571-4771, 4620-4887, 4724-5902
50/1482140CB1/ 2687	1-481, 96-2687, 314-881, 432-1087, 438-1163, 455-1175, 464-1061, 497-1089, 547-1046, 548-1225, 550-1225, 555-1162, 559-1250, 583-1169, 629-1278, 675-1354, 696-1292, 704-1353, 705-1313, 733-1353, 747-1213, 758-1514, 785-1405, 816-1356, 831-1032, 852-1352, 853-1127, 869-1563, 882-1352, 884-1352, 894-1353, 928-1399, 936-1479, 947-1080, 947-1175, 947-1329, 998-1589
51/394992CB1/ 8280	1-193, 1-198, 80-494, 184-738, 184-809, 235-849, 249-732, 271-853, 499-983, 503-1164, 608-1119, 709-1120, 718-1005, 729-1119, 753-1028, 786-1054, 809-1119, 809-1135, 809-1277, 825-1445, 838-1103, 859-1288, 873-1461, 1023-1260, 1032-1249, 1077-1363, 1089-1509, 1089-1615, 1108-1367, 1167-1651, 1167-1658, 1185-1748, 1204-1767, 1262-1883, 1273-1912, 1290-1817, 1349-1904, 1361-1880, 1375-2029, 1407-1898, 1412-1937, 1413-1884, 1451-2089, 1472-2038, 1496-2089, 1497-2028, 1561-1783, 1597-2229, 1639-1866, 1640-2229, 1651-2326, 1679-1906, 1684-1954, 1692-1990, 1704-2336, 1705-1903, 1715-1954, 1717-2294, 1721-2324, 1754-2315, 1812-2370, 1827-2348, 1830-2317, 1901-2370, 1905-2443, 1905-2445, 1909-2339, 1919-2367, 1949-2359, 1969-2363, 1983-2390, 2020-2334, 2022-2381, 2053-2362, 2242-2645, 2365-2866, 2404-2500, 2442-2775, 2443-2582, 2453-2992, 2453-3214, 2585-2631, 2762-3331, 2763-3041, 2856-3300, 2888-3099, 2888-3454, 2928-3776, 2969-3569, 3028-3566, 3028-3582, 3028-3655, 3110-3815, 3112-3861, 3138-3861, 3148-3862, 3194-3862, 3222-3862, 3225-3706, 3271-3841, 3271-3876, 3274-3826, 3304-3876, 3318-3876, 3322-3862, 3331-3876,

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
	3333-3868, 3333-3876, 3350-3394, 3350-3876, 3352-3875, 3361-3876, 3366-3876, 3367-3797, 3370-3876, 3380-3876, 3381-3876, 3386-3876, 3390-3876, 3395-3862, 3396-3876, 3398-3903, 3405-3876, 3409-3863, 3423-3876, 3424-3862, 3456-3758, 3461-3930, 3462-3876, 3752-3932, 3760-4431, 3833-4109, 3838-4102, 3838-4645, 3847-3943, 3980-4571, 3980-4582, 4019-4531, 4105-4533, 4120-4493, 4168-4585, 4232-4617, 4282-4658, 4308-4789, 4408-5143, 4410-5161, 4434-4693, 4434-4705, 4434-4773, 4434-4808, 4434-4949, 4434-5015, 4434-5053, 4440-4903, 4441-4544, 4441-5012, 4441-5053, 4448-5078, 4449-4933, 4461-5015, 4484-5138, 4514-5068, 4521-5139, 4527-5199, 4528-5141, 4553-5139, 4568-4906, 4570-5139, 4570-5141, 4578-5139, 4585-5139, 4585-5141, 4590-5139, 4601-5113, 4611-5139, 4613-5139, 4643-5139, 4651-5139, 4665-5140, 4670-5139, 4676-5139, 4682-5139, 4695-5139, 4713-5139, 4747-5044, 4748-5139, 4753-5139, 4768-5139, 4798-5199, 4834-5092, 4834-5447, 4844-5139, 4925-5378, 4964-5325, 4989-5661, 5047-5222, 5047-5316, 5047-5488, 5076-5325, 5135-5807, 5153-5523, 5160-5801, 5194-6052, 5254-6000, 5292-5832, 5331-5928, 5361-6071, 5394-6234, 5427-5524, 5458-6051, 5465-6142, 5559-5803, 5559-5822, 5559-5830, 5559-5885, 5559-6058, 5559-6184, 5572-6192, 5576-5810, 5576-6097, 5592-5762, 5697-6174, 5710-6343, 5744-5969, 5779-6366, 5813-6437, 5828-6437, 5840-6283, 5923-6494, 5924-6469, 5956-6468, 5971-6541, 5997-6379, 6028-6290, 6028-6563, 6084-6504, 6095-6510, 6101-6711, 6128-6349, 6139-6529, 6167-6307, 6199-6868, 6201-6475, 6218-6529, 6244-6962, 6327-6933, 6391-6955, 6391-6974, 6418-7066, 6454-7073, 6504-7071, 6633-6894, 6641-7055, 6656-7024, 6749-6921, 6816-7025, 6816-7379, 6817-7061, 6842-7119, 6851-7395, 6867-7361, 6885-7290, 6920-7031, 7000-7346, 7019-7382, 7029-7319, 7042-7348, 7052-7615, 7097-7364, 7120-7392, 7159-7403, 7170-7458, 7176-7653, 7179-7653, 7190-7376, 7194-7492, 7245-7374, 7246-7719, 7267-7337, 7562-7662, 7566-7798, 7567-7882, 7570-7624, 7570-7662, 7663-8218, 7663-8236, 7664-7981, 7725-8280, 7744-7836, 7744-8280, 7788-8118, 7795-8228, 7805-8228, 7863-8280, 7876-8126, 7905-8280, 7970-8280, 7978-8271, 8039-8280, 8042-8280, 8048-8280, 8057-8280, 8068-8280, 8070-8280, 8084-8280, 8091-8280, 8093-8280, 8106-8280, 8125-8280, 8130-8280
52/5093550CB1/ 2597	1-167, 1-676, 1-2597, 7-516, 36-688, 62-1919, 200-486, 480-745, 552-589, 552-725, 552-979, 552-1059, 552-1087, 552-1329, 552-1339, 559-589, 559-824, 559-1161, 570-810, 570-1413, 577-1833, 606-1057, 606-1146, 606-1245, 606-1255, 606-1665, 606-1675, 606-1749, 606-1759, 610-894, 610-1497, 610-1507, 649-895, 649-978, 649-1581, 661-816, 661-829, 661-1917, 662-1230, 729-824, 734-890, 745-895, 746-1314, 774-1927, 818-921, 829-979, 830-1398, 858-997, 858-1927, 888-925, 894-979, 902-939, 913-1059, 914-1438, 978-1087, 985-1081, 985-1917, 1004-1566, 1030-1108, 1030-1165, 1030-1927, 1035-1151, 1065-1102, 1065-1161, 1074-1192, 1074-1249, 1074-1917, 1081-1235, 1082-1614, 1140-1270, 1140-1333, 1140-1927, 1146-1255, 1165-1307, 1172-1734, 1233-1339, 1239-1271, 1249-1403, 1250-1774, 1278-1354, 1278-1415, 1278-1927, 1314-1413, 1315-1501, 1315-1917, 1316-1425, 1333-1487, 1334-1895, 1392-1522, 1392-1583, 1392-1927, 1398-1507, 1400-1509, 1417-1571, 1418-1927, 1476-1669, 1476-1903, 1482-1581, 1484-1593, 1501-1655, 1502-1917, 1560-1597, 1566-1675, 1568-1622, 1585-1739, 1586-1927, 1650-1759, 1651-1775, 1651-1835, 1651-1916, 1652-1775, 1657-1753, 1657-1903, 1669-1823, 1670-1903, 1734-1833, 1736-1774, 1753-1908, 1754-1903, 1812-1849, 1818-2157, 1818-2384, 1818-2428, 1818-2439,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1818-2588, 1820-1866, 1838-1916, 1933-1986, 1933-2023, 2150-2189, 2150-2190, 2150-2197, 2150-2204, 2150-2260, 2150-2273, 2150-2274, 2151-2189, 2151-2190, 2157-2194, 2157-2260, 2157-2313, 2158-2190, 2433-2522
53/748797/CB1/ 725	1-296, 18-725
54/1706514/CB1/ 1952	1-273, 1-596, 1-1939, 9-477, 9-606, 9-668, 24-278, 91-389, 227-477, 227-718, 227-983, 243-891, 420-985, 439-664, 439-903, 449-952, 536-975, 546-1221, 613-1221, 615-894, 627-1150, 643-1292, 644-1227, 670-1311, 679-1312, 745-1188, 808-955, 810-958, 837-955, 854-1196, 854-1220, 854-1282, 878-1421, 887-1425, 892-1039, 899-1039, 907-954, 907-958, 907-1124, 907-1141, 907-1152, 907-1217, 907-1253, 907-1457, 913-958, 914-1385, 924-1257, 924-1366, 970-1039, 971-1039, 997-1247, 997-1282, 998-1039, 1060-1320, 1067-1534, 1075-1318, 1075-1356, 1075-1389, 1075-1395, 1075-1460, 1075-1477, 1075-1583, 1075-1625, 1090-1388, 1092-1553, 1092-1554, 1094-1206, 1146-1198, 1159-1210, 1161-1198, 1165-1415, 1165-1457, 1166-1210, 1183-1568, 1190-1486, 1190-1524, 1190-1554, 1190-1556, 1190-1583, 1190-1618, 1230-1378, 1251-1618, 1258-1488, 1258-1625, 1262-1374, 1314-1364, 1314-1365, 1323-1618, 1327-1378, 1329-1364, 1333-1629, 1333-1752, 1342-1386, 1351-1625, 1377-1457, 1426-1618, 1430-1542, 1466-1921, 1482-1533, 1482-1605, 1487-1618, 1495-1546, 1501-1625, 1519-1618, 1545-1835, 1566-1596, 1566-1613, 1566-1616, 1566-1618, 1566-1677, 1566-1689, 1651-1689, 1730-1905, 1730-1952, 1731-1952
55/7488247/CB1/ 1213	1-164, 4-200, 4-203, 121-690, 433-834, 436-969, 442-759, 442-880, 442-1022, 482-775, 483-756, 490-802, 509-1139, 559-1213, 604-1213, 617-1198, 740-823, 745-1212, 753-886, 771-1211, 773-930, 799-1039, 799-1074, 808-1001, 812-1212, 845-1026, 851-987, 851-1109, 905-1028, 905-1130, 905-1139, 905-1166, 905-1167, 905-1176, 905-1206, 905-1213, 909-1132, 909-1198, 909-1213, 913-1150, 921-1191, 933-1151, 957-1213, 961-1206, 967-1206, 974-1206, 978-1213, 986-1206, 998-1213, 1003-1213, 1014-1213, 1015-1213, 1018-1213, 1035-1193, 1036-1213, 1040-1213, 1045-1206, 1059-1206, 1069-1213, 1072-1213, 1077-1167, 1078-1213, 1082-1207, 1084-1207, 1100-1213, 1103-1213, 1110-1213, 1129-1206, 1143-1211
56/1427269/CB1/ 2257	1-801, 3-605, 4-467, 14-512, 14-528, 14-550, 14-576, 14-593, 14-604, 14-605, 14-649, 16-304, 72-666, 73-463, 73-560, 88-617, 108-338, 108-394, 108-454, 108-541, 108-564, 108-575, 108-606, 108-620, 108-646, 108-657, 128-824, 141-817, 151-379, 162-362, 169-803, 169-824, 171-728, 189-485, 218-773, 265-728, 271-781, 285-993, 290-763, 315-605, 316-856, 321-729, 334-869, 379-1050, 383-869, 383-984, 429-881, 448-1133, 457-873, 467-862, 493-863, 504-1084, 546-1159, 579-1245, 588-1127, 588-1186, 635-1139, 658-1325, 714-1298, 718-1344, 730-1225, 802-1349, 931-1188, 1151-1729, 1281-1848, 1309-1853, 1353-1595, 1360-1970, 1360-1986, 1374-1538, 1380-1941, 1388-1537, 1389-1914, 1441-1940, 1448-2027, 1462-1685, 1463-1732, 1483-2056, 1484-2109, 1484-2140, 1526-1860, 1553-2080, 1564-2094, 1583-2257, 1595-1841, 1597-2090

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
57/103135CB1/ 2415	1-298, 63-336, 113-465, 118-409, 132-474, 138-377, 138-618, 138-912, 138-942, 157-931, 166-599, 166-664, 262-763, 351-967, 359-1116, 364-1011, 440-970, 445-1122, 474-640, 509-1026, 513-1035, 522-1128, 524-1212, 529-1219, 565-1236, 601-1303, 605-1250, 613-1143, 613-1300, 636-1135, 656-1191, 658-1191, 695-1381, 697-1299, 745-972, 757-924, 759-949, 774-1242, 801-1499, 818-1354, 823-1457, 826-939, 827-1373, 833-1535, 834-1527, 855-1471, 876-1273, 909-1376, 933-1402, 940-1608, 981-1376, 989-1535, 998-1196, 1009-1455, 1011-1392, 1027-1554, 1037-1474, 1047-1544, 1081-1558, 1085-1712, 1085-1724, 1097-1410, 1097-1719, 1106-1761, 1137-1698, 1205-1789, 1209-1790, 1209-1820, 1276-1814, 1281-1820, 1286-1926, 1343-1917, 1355-2137, 1356-2118, 1382-1836, 1393-1667, 1446-2056, 1447-1681, 1514-1968, 1525-1803, 1525-2039, 1543-1974, 1560-2121, 1585-2057, 1588-1829, 1599-2067, 1609-2120, 1617-2160, 1696-2132, 1743-2161, 1743-2415, 1749-2174, 1785-2188, 1791-2184, 1806-2132, 1808-2184, 1882-2415, 1899-2109, 1903-2090, 1903-2148, 1903-2161, 1955-2132, 1955-2161, 1996-2189
58/1907346CB1/ 4272	1-784, 367-816, 367-4272, 413-1008, 483-1024, 487-830, 548-1186, 558-944, 691-1367, 723-1464, 779-1260, 779-1678, 787-1406, 842-1381, 842-1452, 842-1494, 916-1550, 928-1270, 979-1644, 979-1679, 981-1546, 1115-1144, 1115-1195, 1115-1210, 1115-1214, 1115-1228, 1115-1284, 1115-1286, 1115-1296, 1115-1305, 1115-1308, 1115-1313, 1115-1324, 1115-1326, 1115-1378, 1115-1381, 1115-1393, 1115-1402, 1115-1428, 1115-1476, 1115-1480, 1115-1483, 1115-1486, 1115-1558, 1115-1570, 1115-1729, 1115-1740, 1115-2302, 1115-2468, 1115-2637, 1116-1145, 1116-1168, 1116-1211, 1116-1213, 1116-1347, 1116-1404, 1116-1478, 1116-1494, 1116-1505, 1116-1531, 1116-1633, 1116-1656, 1116-1714, 1116-1732, 1116-1882, 1116-2221, 1117-1378, 1118-1213, 1119-1363, 1119-1546, 1121-1213, 1125-1378, 1126-1477, 1129-1213, 1130-1404, 1135-1333, 1137-1488, 1138-1754, 1149-1546, 1155-1213, 1167-1213, 1167-1699, 1167-1740, 1171-1213, 1182-1378, 1183-1714, 1184-1378, 1184-1393, 1187-1714, 1188-1213, 1188-1372, 1192-1662, 1193-1297, 1193-1387, 1193-1388, 1193-1867, 1193-2050, 1194-1897, 1197-1378, 1203-1591, 1218-1378, 1227-1471, 1227-1546, 1243-1387, 1244-1714, 1245-1378, 1246-1378, 1260-1546, 1260-1656, 1266-1714, 1272-1309, 1272-1393, 1272-1573, 1272-1654, 1272-1732, 1272-2075, 1272-2317,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1275-1480, 1275-1488, 1275-2074, 1278-1393, 1278-1404, 1278-1546, 1283-1549, 1283-1573, 1283-1620, 1283-1632, 1283-1654, 1283-1714, 1283-1717, 1283-1738, 1283-1783, 1283-1807, 1283-2805, 1284-1535, 1287-1648, 1294-1740, 1305-1658, 1306-1908, 1313-1714, 1321-1378, 1333-1546, 1335-1378, 1335-1807, 1335-1867, 1337-1378, 1337-1882, 1344-1998, 1349-1882, 1352-1404, 1352-1730, 1355-1882, 1356-1392, 1356-1393, 1356-1450, 1356-1456, 1356-1477, 1356-1564, 1356-1656, 1356-1657, 1356-1738, 1356-1756, 1356-2401, 1361-1810, 1362-2065, 1365-1546, 1371-1764, 1374-1546, 1391-1639, 1391-1714, 1395-1882, 1410-1882, 1418-1714, 1419-1546, 1425-1714, 1428-1810, 1436-1732, 1438-1740, 1439-1477, 1439-1855, 1440-1882, 1451-1584, 1451-1601, 1451-1717, 1451-1738, 1451-1784, 1451-1800, 1451-1801, 1451-1882, 1451-1885, 1451-1945, 1451-1980, 1451-1985, 1451-2804, 1460-1714, 1461-1703, 1462-1732, 1473-1824, 1474-2065, 1483-1882, 1489-2035, 1511-2287, 1513-1633, 1518-1714, 1519-2050, 1520-1564, 1520-1633, 1520-1740, 1523-2050, 1524-1630, 1524-1633, 1527-1732, 1527-1906, 1527-1981, 1527-2149, 1527-2326, 1528-1563, 1529-1984, 1530-2075, 1533-1714, 1538-1882, 1539-1920, 1554-1714, 1563-1807, 1563-1882, 1563-2050, 1567-1714, 1579-1633, 1595-1882, 1596-1982, 1603-1714, 1604-1630, 1604-1658, 1604-1740, 1604-1817, 1605-2050, 1608-1654, 1608-1732, 1609-1654, 1611-1740, 1611-1807, 1612-2019, 1612-2075, 1619-1761, 1619-1774, 1619-1885, 1619-1906, 1619-1968, 1619-1985, 1619-2050, 1619-2053, 1619-2074, 1619-2806, 1620-1871, 1623-1985, 1665-2097, 1676-2075, 1676-2161, 1680-2416, 1686-1882, 1690-2221, 1691-1992, 1691-2050, 1691-2218, 1691-2221, 1692-1732, 1692-1740, 1692-1807, 1692-2074, 1692-2158, 1692-2317, 1692-2495, 1695-1741, 1697-2161, 1701-1882, 1701-2311, 1704-1731, 1704-1732, 1704-1756, 1704-1966, 1704-1981, 1704-2137, 1704-2203, 1704-2227, 1704-2230, 1704-2809, 1707-2065, 1710-1882, 1725-2221, 1727-1975, 1727-2050, 1731-2221, 1755-1882, 1761-2050, 1764-2075, 1770-2221, 1787-2053, 1787-2069, 1787-2074, 1787-2218, 1787-2236, 1787-2412, 1788-2074, 1796-2050, 1816-2074, 1817-2221, 1821-2271, 1822-2344, 1825-2074, 1832-1882, 1838-2626, 1839-1882, 1841-1882, 1842-2050, 1845-2075, 1855-2075, 1856-1897, 1856-1909, 1856-1969, 1856-2234, 1856-2287, 1856-2317, 1860-1896, 1860-1908, 1860-1909, 1860-2075, 1860-2134, 1860-2250, 1860-2342, 1860-2485, 1863-1966, 1863-1969, 1864-1969, 1865-2329, 1866-2412, 1869-2050, 1873-2050, 1876-2237, 1878-2050, 1887-2221, 1892-2075, 1895-2074, 1895-2221, 1897-2075, 1898-2075, 1917-2050, 1918-2050,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1929-1981, 1929-2049, 1932-2317, 1940-1969, 1940-2468, 1941-2075, 1944-1982, 1944-2437, 1944-2468, 1948-1984, 1949-2122, 1949-2137, 1949-2161, 1949-2206, 1949-2220, 1949-2237, 1949-2302, 1949-2305, 1949-2345, 1949-2400, 1949-2405, 1949-2410, 1949-2806, 1950-2653, 1956-2069, 1956-2218, 1956-2227, 1956-2296, 1956-2317, 1956-2572, 1966-2448, 1978-2569, 1988-2075, 1993-2317, 2016-2065, 2016-2119, 2016-2227, 2016-2248, 2016-2302, 2016-2352, 2016-2402, 2023-2074, 2023-2221, 2024-2074, 2024-2401, 2027-2074, 2028-2464, 2028-2468, 2031-2134, 2032-2075, 2032-2134, 2032-2162, 2040-2082, 2040-2184, 2040-2190, 2040-2291, 2040-2302, 2040-2305, 2040-2329, 2040-2374, 2040-2388, 2040-2416, 2040-2468, 2040-2568, 2040-2573, 2040-2806, 2043-2074, 2057-2391, 2070-2412, 2070-2468, 2074-2625, 2090-2326, 2091-2662, 2106-2302, 2110-2302, 2110-2405, 2110-2637, 2111-2637, 2112-2148, 2112-2150, 2112-2151, 2112-2317, 2112-2494, 2112-2572, 2112-2746, 2115-2151, 2115-2221, 2117-2580, 2118-2148, 2118-2806, 2123-2386, 2123-2391, 2123-2401, 2124-2302, 2124-2317, 2127-2520, 2128-2302, 2130-2302, 2153-2393, 2153-2468, 2168-2637, 2169-2302, 2170-2302, 2174-2637, 2183-2605, 2184-2468, 2184-2580, 2192-2217, 2192-2221, 2192-2317, 2192-2401, 2196-2218, 2196-2317, 2196-2401, 2196-2637, 2199-2236, 2201-2236, 2208-2439, 2208-2468, 2208-2494, 2208-2525, 2208-2556, 2208-2586, 2208-2637, 2208-2641, 2208-2707, 2208-2726, 2208-2736, 2208-2806, 2210-2468, 2211-2573, 2245-2302, 2250-2806, 2253-2326, 2253-2578, 2253-2637, 2257-2468, 2259-2302, 2259-2495, 2259-2746, 2261-2302, 2261-2894, 2268-2804, 2273-2914, 2276-2317, 2276-2328, 2276-2389, 2276-2401, 2279-2919, 2280-2316, 2280-2317, 2280-2328, 2280-2401, 2283-2317, 2283-2389, 2285-2750, 2286-2806, 2289-2468, 2292-2389, 2292-2569, 2292-2580, 2292-2809, 2295-2662, 2298-2468, 2312-2935, 2315-2563, 2315-2637, 2317-2887, 2318-2988, 2342-2637, 2343-2468, 2349-2637, 2358-3004, 2362-2412, 2364-2389, 2364-2401, 2364-2403, 2364-2580, 2364-2738, 2364-2776, 2365-2935, 2369-2389, 2379-2627, 2379-2641, 2379-2665, 2379-2713, 2379-2741, 2379-2803, 2379-2805, 2379-2806, 2383-2637, 2395-2495, 2397-2754, 2398-2806, 2405-2946, 2421-2804, 2424-2637, 2427-2468, 2429-2468, 2447-2805, 2454-2580, 2454-2806, 2457-2637, 2457-2809, 2475-2805, 2475-2806, 2478-2637, 2480-2805, 2483-2733, 2483-3027, 2485-2805, 2486-2805, 2505-2637, 2505-2862, 2506-2637, 2516-3149, 2520-2804, 2520-2806, 2528-2569, 2528-2580, 2529-2805, 2532-2571, 2543-2685, 2543-2806, 2544-2795, 2554-2809, 2565-2805, 2566-2806, 2573-2805,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	2589-3078, 2589-3095, 2589-3254, 2599-2804, 2610-2806, 2611-2804, 2612-2666, 2615-2804, 2621-2806, 2622-3063, 2623-2835, 2625-2806, 2631-2806, 2634-2899, 2651-2805, 2659-3328, 2673-2746, 2674-3326, 2679-2804, 2685-2805, 2696-2807, 2696-2809, 2697-2804, 2703-2749, 2703-2809, 2704-2806, 2705-2809, 2705-2892, 2706-2892, 2717-2809, 2719-2746, 2733-2804, 2734-2809, 2741-2804, 2749-2965, 2758-3123, 2759-3029, 2763-2806, 2780-2809, 2797-3072, 2801-3098, 2810-3199, 2820-3438, 2861-2899, 2871-3405, 2881-3546, 2884-3129, 2926-3718, 2935-3545, 2959-3218, 2959-3254, 2974-3272, 2974-3487, 2994-3255, 3081-3277, 3085-3355, 3085-3667, 3087-3298, 3090-3665, 3090-3810, 3134-3635, 3176-3442, 3201-3790, 3202-3444, 3202-3461, 3202-3658, 3202-3809, 3203-3477, 3203-3809, 3214-3511, 3215-3699, 3221-3727, 3221-3770, 3221-3845, 3230-3809, 3244-3806, 3246-3621, 3297-3820, 3297-3884, 3297-3999, 3298-3611, 3310-3692, 3311-3570, 3328-3621, 3364-4201, 3373-3634, 3380-4066, 3382-3680, 3407-3806, 3448-3790, 3448-3844, 3452-3792, 3466-3807, 3468-3792, 3470-3675, 3477-3741, 3484-4126, 3485-3781, 3491-3790, 3498-4201, 3499-3796, 3504-3841, 3515-3802, 3524-3790, 3531-3726, 3545-3789, 3579-3790, 3596-3869, 3828-3862, 3942-3996, 3962-3990, 3962-3993, 3962-3996
59/3041036CB1/ 1823	1-549, 1-623, 121-333, 203-1823, 223-333, 249-333, 739-942, 739-1210, 741-1275, 743-1123, 778-1695, 807-1614, 851-1362, 862-1779, 869-1499, 879-1022, 879-1445, 975-1782, 1121-1751, 1215-1781, 1245-1779, 1271-1782, 1327-1798, 1415-1795, 1579-1782, 1635-1778
60/3856879CB1/ 2253	1-506, 43-2223, 115-599, 140-336, 361-1040, 467-1011, 503-1095, 675-1361, 680-1123, 700-948, 811-970, 811-1301, 811-1552, 814-1014, 817-858, 817-870, 817-881, 817-885, 817-896, 817-941, 817-953, 817-956, 817-1374, 818-885, 818-953, 819-1469, 823-885, 824-953, 831-893, 831-941, 832-882, 832-884, 832-892, 832-962, 841-881, 842-941, 860-1260, 868-941, 869-941, 909-959, 909-962, 909-1048, 909-1105, 909-1302, 909-1507, 916-940, 916-941, 916-953, 916-976, 916-1046, 936-1600, 937-1225, 958-1251, 972-1248, 998-1614, 1005-1673, 1033-1110, 1036-1088, 1063-1088, 1063-1109, 1063-1136, 1063-1222, 1063-1553, 1063-1804, 1064-1088, 1064-1136, 1071-1553, 1073-1136, 1118-1432, 1148-1796, 1156-1184, 1156-1217, 1156-1302, 1156-1441, 1156-1507, 1156-1520, 1156-1522, 1164-1467, 1168-1217, 1168-1374, 1168-1385, 1168-1467, 1184-1710, 1184-1830, 1191-1374, 1228-1467, 1235-1698, 1237-1271, 1237-1290, 1237-1300, 1237-1301, 1237-1302, 1237-1368, 1237-1373, 1237-1387, 1237-1457, 1237-1543, 1237-1552, 1237-1794, 1239-1328, 1239-1361, 1239-1362, 1239-1373, 1239-1374, 1240-1299, 1240-1301, 1240-1394, 1240-1467, 1240-1606, 1244-1302, 1245-1374, 1245-1424, 1245-1441, 1245-1638, 1245-1843, 1252-1302, 1259-1957, 1282-1441, 1285-1373, 1288-1374, 1289-1374, 1308-1374, 1314-1385,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1336-1374, 1336-1383, 1336-1385, 1336-1388, 1336-1441, 1336-1497, 1336-1498, 1336-1507, 1336-1529, 1336-1541, 1336-1542, 1336-1553, 1336-1606, 1336-1635, 1341-1394, 1342-1467, 1345-1374, 1345-1507, 1359-1385, 1359-1635, 1396-2036, 1408-1507, 1408-1520, 1408-1635, 1412-1530, 1412-1541, 1453-1530, 1456-1529, 1456-1552, 1457-1552, 1462-2145, 1478-1909, 1482-1635, 1486-1552, 1489-1553, 1489-1638, 1489-1804, 1490-1804, 1492-1519, 1492-1520, 1492-1638, 1492-1843, 1497-1541, 1502-1553, 1510-1762, 1510-1984, 1513-1552, 1515-2204, 1521-2171, 1545-1784, 1545-1785, 1573-1638, 1573-2126, 1579-2184, 1580-1638, 1580-1833, 1580-2143, 1581-1606, 1581-1634, 1625-2214, 1659-1698, 1659-1748, 1659-1781, 1659-1782, 1659-1794, 1671-1781, 1672-1722, 1672-1843, 1681-2152, 1698-2210, 1708-1781, 1708-1804, 1711-2032, 1714-1794, 1718-2194, 1728-1986, 1728-2238, 1735-1760, 1735-1781, 1735-1808, 1736-2018, 1738-1774, 1738-1804, 1741-1804, 1741-1809, 1743-1782, 1743-1809, 1744-1843, 1760-2232, 1765-1804, 1770-2214, 1791-2220, 1798-2224, 1799-2253, 1817-2218, 1823-2220, 1868-2218, 1887-2142, 1887-2216, 1887-2245, 1915-2204, 1947-2219, 2019-2253, 2084-2253
61/4178665CBI/ 2788	1-609, 1-612, 5-794, 248-686, 275-557, 292-800, 374-970, 388-988, 412-912, 418-1026, 455-718, 455-998, 525-626, 525-711, 537-600, 537-694, 537-884, 537-889, 562-630, 562-632, 562-694, 567-722, 569-636, 569-714, 569-774, 569-883, 570-711, 570-715, 570-722, 570-794, 578-604, 578-628, 578-630, 578-640, 578-690, 578-700, 578-711, 578-712, 578-722, 578-759, 578-774, 578-787, 578-794, 578-805, 578-814, 578-879, 578-1080, 578-1131, 579-1124, 580-879, 581-722, 581-975, 586-698, 586-719, 586-721, 586-759, 586-764, 586-882, 586-1056, 586-1058, 587-722, 588-794, 621-688, 621-766, 621-778, 621-973, 645-890, 649-879, 649-1124, 651-722, 651-1067, 654-722, 662-691, 662-722, 662-759, 662-787, 662-882, 662-883, 662-975, 664-1047, 665-1061, 672-722, 673-876, 673-888, 673-968, 673-1131, 686-722, 686-956, 686-975, 693-854, 693-879, 693-967, 693-1135, 693-1146, 693-1160, 693-1194, 697-879, 735-883, 735-1134, 745-879, 745-975, 745-1124, 747-1377, 749-882, 819-1040, 819-1050, 819-1393, 821-879, 821-881, 821-882, 821-883, 821-889, 821-966, 821-976, 821-1026, 821-1046, 821-1120, 821-1132, 821-1134, 821-1135, 821-1393, 824-964, 824-1049, 824-1061, 828-883, 828-888,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	830-956, 836-883, 836-950, 836-966, 836-1122, 836-1131, 836-1134, 838-965, 838-1059, 872-1142, 892-1135, 892-1383, 902-1525, 906-1134, 911-1047, 911-1124, 912-972, 914-963, 914-968, 914-1019, 914-1054, 914-1058, 914-1135, 920-956, 920-1034, 920-1057, 920-1059, 920-1067, 920-1387, 922-976, 922-1044, 922-1055, 922-1083, 922-1134, 922-1478, 938-1134, 960-1142, 983-1393, 986-1058, 987-1232, 987-1310, 987-1588, 997-1124, 998-1028, 998-1041, 998-1043, 998-1047, 998-1066, 998-1131, 998-1132, 998-1134, 998-1386, 998-1449, 999-1386, 1001-1058, 1001-1122, 1001-1393, 1002-1036, 1006-1050, 1006-1133, 1006-1134, 1006-1142, 1008-1057, 1029-1558, 1037-1561, 1043-1131, 1048-1317, 1048-1581, 1078-1302, 1088-1142, 1088-1379, 1093-1310, 1106-1275, 1106-1381, 1106-1393, 1114-1300, 1128-1370, 1128-1779, 1138-1573, 1147-1479, 1147-1820, 1149-1766, 1154-1645, 1160-1424, 1174-1228, 1174-1393, 1201-1393, 1256-1294, 1256-1305, 1256-1310, 1256-1381, 1256-1393, 1256-1403, 1256-1456, 1256-1478, 1256-1551, 1256-1636, 1256-1646, 1258-1308, 1258-1311, 1258-1355, 1258-1356, 1258-1384, 1258-1386, 1258-1393, 1258-1447, 1258-1449, 1258-1562, 1258-1638, 1258-1648, 1258-1698, 1258-1814, 1259-1403, 1260-1644, 1274-1553, 1274-1562, 1289-1386, 1306-2022, 1324-1383, 1324-1386, 1324-1393, 1324-1731, 1325-1383, 1325-1393, 1328-1392, 1328-1393, 1328-1562, 1331-1393, 1331-1813, 1333-1600, 1333-1851, 1334-1390, 1334-1393, 1334-1403, 1334-1474, 1334-1546, 1334-1550, 1334-1554, 1334-1635, 1334-1898, 1337-1393, 1340-1454, 1340-1487, 1340-1562, 1340-1639, 1340-1650, 1340-1734, 1340-1812, 1340-1814, 1344-1372, 1344-1456, 1379-1614, 1387-1687, 1405-1478, 1405-1635, 1410-1555, 1410-1562, 1410-1634, 1410-1806, 1410-1815, 1418-1551, 1418-1869, 1423-1479, 1426-1562, 1428-1935, 1462-1813, 1463-1699, 1467-1551, 1469-1551, 1469-1869, 1480-2120, 1492-1520, 1492-1531, 1492-1533, 1492-1551, 1492-1561, 1492-1726, 1492-1736, 1492-1737, 1492-1896, 1492-1898, 1493-1806, 1496-1569, 1496-1636, 1496-1717, 1496-1731, 1496-1989, 1497-2060, 1499-1851, 1505-1562, 1505-1618, 1505-1646, 1505-1803, 1505-1856, 1505-1880, 1505-1897, 1505-1899, 1506-1734, 1508-1645, 1508-1731, 1508-1812, 1508-1814, 1518-1794, 1518-1851, 1520-1731, 1538-1720, 1583-1734, 1583-1831, 1583-1851, 1583-1935, 1586-1803, 1591-1851, 1594-1721, 1594-1729, 1594-1878, 1594-1896, 1594-1898, 1594-1935, 1599-1889, 1599-2151, 1599-2227, 1610-1806, 1610-1867, 1610-1885, 1610-1897, 1611-1806, 1612-1897, 1618-1804, 1632-1869, 1632-1897, 1633-1923, 1633-2259, 1660-2078, 1660-2109, 1664-1692,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1664-1706, 1664-1815, 1664-1860, 1664-1866, 1664-1894, 1664-1899, 1664-1901, 1667-1732, 1667-1880, 1670-1723, 1670-1730, 1670-1778, 1670-1802, 1670-1806, 1670-1807, 1670-1866, 1670-1876, 1670-1877, 1670-1879, 1670-1890, 1671-1899, 1673-1899, 1678-1785, 1678-1805, 1678-1813, 1678-1863, 1678-1869, 1678-1885, 1678-1897, 1688-1935, 1704-1783, 1706-1878, 1715-1880, 1715-1897, 1752-1898, 1753-1935, 1754-1862, 1754-1876, 1754-1879, 1754-1880, 1754-1931, 1754-1933, 1754-1935, 1778-1814, 1778-1899, 1786-1899, 1797-1880, 1799-1898, 1805-1880, 1827-1898, 1832-1899, 1838-1899, 1838-1935, 1839-1879, 1841-1898, 1877-2069, 1886-2111, 1886-2420, 1891-2423, 1953-2607, 1968-2216, 2006-2422, 2007-2330, 2010-2523, 2011-2306, 2011-2576, 2020-2721, 2055-2741, 2061-2682, 2070-2788, 2082-2771, 2095-2724, 2101-2583, 2119-2405, 2119-2595, 2152-2244, 2152-2418, 2193-2369, 2197-2788, 2211-2788, 2246-2774, 2256-2550, 2263-2739, 2297-2722, 2301-2764, 2301-2788, 2335-2466, 2357-2786, 2363-2785, 2363-2788, 2364-2466, 2407-2701, 2407-2704, 2407-2723, 2407-2724, 2407-2785, 2407-2788, 2409-2704
62/7493326CBI/ 3041	1-551, 1-793, 1-1495, 17-198, 83-793, 175-509, 286-509, 338-603, 338-985, 542-2428, 574-1238, 744-1509, 751-1104, 828-1337, 834-1396, 884-1472, 904-942, 904-943, 922-1493, 925-1493, 926-1493, 927-1029, 927-1070, 927-1100, 927-1246, 927-1866, 941-1493, 952-1027, 960-1164, 960-1279, 960-1352, 968-1493, 972-993, 992-1493, 995-1493, 1005-1493, 1026-1493, 1052-1167, 1052-1204, 1052-1248, 1052-1279, 1052-1325, 1052-1326, 1052-1330, 1052-1404, 1052-1411, 1052-1414, 1052-2024, 1052-2111, 1059-1326, 1061-1245, 1069-1509, 1083-1117, 1084-1493, 1093-1493, 1109-1493, 1124-1493, 1127-1426, 1127-1530, 1127-1604, 1135-1244, 1135-1272, 1135-1341, 1135-1498, 1135-1615, 1135-1678, 1135-1687, 1135-1794, 1135-1830, 1136-1448, 1136-1530, 1136-1577, 1136-1579, 1136-2195, 1144-1258, 1144-1364, 1144-1413, 1144-1525, 1172-1325, 1179-1201, 1179-1352, 1179-1577, 1179-1589, 1179-1595, 1179-1615, 1179-1709, 1179-1740, 1179-1747, 1179-1763, 1179-2376, 1209-1272, 1214-1740, 1220-1287, 1220-1341, 1220-1365, 1220-1427, 1220-1457, 1220-1498, 1220-1520, 1220-1530, 1220-1588, 1220-1625, 1220-1661, 1220-1663, 1220-1666, 1220-1711, 1220-1762, 1220-1866, 1220-1954, 1220-2195, 1224-1497, 1227-1362, 1228-1363, 1228-1404, 1231-1709, 1240-1279, 1240-1358, 1250-1740, 1257-1740, 1291-1498,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1291-1845, 1291-1952, 1295-1661, 1295-1666, 1312-1577, 1312-1592, 1312-1595, 1340-1493, 1355-1924, 1355-2034, 1378-1449, 1386-1595, 1387-1702, 1394-1665, 1438-2014, 1439-2136, 1440-1679, 1462-1530, 1471-1672, 1471-1700, 1471-1803, 1471-1829, 1471-1830, 1471-1834, 1471-2169, 1472-1709, 1472-1740, 1472-1763, 1472-1866, 1480-1755, 1508-1661, 1523-2101, 1523-2195, 1555-1615, 1555-1708, 1560-1830, 1605-2172, 1610-1845, 1610-2195, 1639-1782, 1639-1866, 1639-1951, 1639-1953, 1639-1999, 1639-2002, 1639-2194, 1640-1816, 1640-1845, 1640-1866, 1640-1913, 1640-1915, 1640-1918, 1640-2034, 1640-2453, 1646-1866, 1682-1829, 1715-1925, 1715-1709, 1715-2034, 1715-2195, 1723-1786, 1723-1867, 1723-1931, 1723-2195, 1730-2001, 1765-1913, 1775-2169, 1776-2015, 1800-1940, 1800-2034, 1800-2169, 1800-2195, 1808-2034, 1808-2195, 1814-2080, 1815-1950, 1839-2195, 1844-1953, 1852-2195, 1883-2195, 1891-1940, 1891-2030, 1891-2034, 1891-2102, 1891-2436, 1891-2453, 1892-2008, 1892-2045, 1892-2132, 1892-2168, 1892-2170, 1892-2172, 1892-2195, 1896-2183, 1900-2025, 1900-2123, 1900-2172, 1900-2195, 1900-2255, 1900-2284, 1910-1947, 1933-2034, 1975-2172, 1975-2194, 1985-2172, 1985-2195, 1987-2034, 1988-2030, 1996-2034, 2015-2168, 2060-2195, 2067-2195, 2068-2195, 2069-2172, 2084-2122, 2104-2195, 2112-2195, 2117-2195, 2146-2207, 2278-2587, 2303-2900, 2303-3041, 2305-2341, 2305-2373, 2305-2376, 2314-2374, 2314-2375, 2314-2451, 2314-2453, 2315-2382, 2315-2421, 2315-2453, 2322-2453, 2326-2373, 2326-2377, 2374-2843, 2846-2867
63/1553836CBI/ 3445	1-359, 1-503, 1-529, 1-575, 6-214, 6-593, 13-676, 18-684, 19-584, 28-612, 29-763, 35-703, 39-877, 41-673, 41-893, 158-918, 180-398, 213-497, 216-850, 272-417, 326-601, 326-857, 327-1063, 330-761, 393-634, 393-960, 394-1074, 437-955, 533-1034, 565-773, 565-1158, 574-1177, 591-1170, 616-1177, 622-914, 683-922, 688-1176, 688-1210, 706-916, 740-1319, 861-1433, 861-1443, 892-1564, 969-1572, 1055-1366, 1083-1594, 1117-2702, 1137-1810, 1167-1483, 1206-1733, 1220-1840, 1254-1933, 1270-1768, 1282-1422, 1306-1464, 1306-1556, 1320-1584, 1320-1593, 1320-1842, 1320-1862, 1320-2068, 1382-1614, 1382-1673, 1387-1650, 1393-1666, 1414-2195, 1437-2061, 1450-1739, 1453-1732, 1460-2055, 1466-1741, 1478-2030, 1482-1952, 1482-1957, 1482-2008, 1483-1845, 1500-1785, 1506-2107, 1561-1999, 1609-2201, 1632-2219, 1645-1875, 1649-1982, 1649-1984, 1665-2294, 1694-2298, 1703-1907, 1713-2252, 1725-2170, 1725-2488, 1726-1988, 1728-2007, 1733-2012, 1733-2314, 1760-2045, 1762-2002, 1769-2191, 1770-2050, 1783-2382, 1798-2543, 1802-2493, 1847-2133, 1864-2087, 1866-2056, 1896-2163, 1903-2610, 1915-2676, 1917-2652, 1928-2059, 1935-2238, 1937-2587, 1938-2567, 1967-2363, 1988-2640, 1989-2609, 2002-2230, 2002-2640, 2003-2248, 2003-2258, 2021-2551, 2053-2636, 2072-2716, 2084-2705, 2087-2378,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	2090-2666, 2105-2745, 2110-2368, 2115-2344, 2115-2354, 2115-2687, 2115-2708, 2118-2334, 2118-2593, 2120-2801, 2121-2745, 2121-2792, 2137-2747, 2137-2932, 2138-2683, 2141-2432, 2143-2354, 2145-2495, 2152-2792, 2167-2367, 2182-2456, 2195-2390, 2204-2470, 2209-2876, 2216-2836, 2217-2782, 2223-2744, 2223-2883, 2224-2436, 2224-2806, 2225-2794, 2253-2406, 2257-2527, 2258-2497, 2266-2778, 2273-2938, 2275-2989, 2278-2533, 2289-2462, 2300-2966, 2305-2688, 2316-2874, 2316-2932, 2323-2641, 2324-2655, 2324-2932, 2337-2628, 2343-2604, 2344-2867, 2349-2638, 2360-2913, 2361-2932, 2362-2914, 2363-2913, 2369-2619, 2370-2948, 2371-2687, 2382-2678, 2387-2932, 2389-2932, 2395-2629, 2396-2652, 2396-2698, 2401-2659, 2401-2688, 2402-2646, 2409-2953, 2409-3151, 2418-2669, 2418-2670, 2419-2990, 2429-3098, 2431-3111, 2433-3249, 2439-3066, 2441-2724, 2447-2575, 2448-2989, 2451-3253, 2454-2929, 2476-2915, 2486-2739, 2512-2758, 2516-2792, 2519-2785, 2523-2815, 2524-2980, 2554-2586, 2554-2989, 2559-2806, 2574-3105, 2574-3208, 2579-3316, 2581-2843, 2581-3186, 2591-3245, 2625-3309, 2631-3248, 2638-2866, 2638-3313, 2643-3306, 2651-2867, 2652-3265, 2653-3323, 2677-2936, 2678-3309, 2682-3284, 2683-2974, 2683-3306, 2688-3112, 2690-3266, 2700-2972, 2712-3323, 2717-3323, 2720-2937, 2720-3197, 2724-2978, 2731-2921, 2731-3325, 2740-3323, 2743-3257, 2745-3137, 2751-3382, 2753-3043, 2757-3044, 2760-2960, 2769-3045, 2778-3164, 2782-3015, 2788-3072, 2792-3323, 2806-3042, 2810-3323, 2821-3325, 2827-3118, 2827-3276, 2827-3277, 2843-3044, 2847-3323, 2848-3320, 2849-3321, 2854-3312, 2854-3323, 2855-3125, 2860-3317, 2861-3138, 2866-3175, 2866-3325, 2867-3323, 2870-3320, 2877-3319, 2889-3320, 2894-3314, 2902-3317, 2902-3320, 2903-3320, 2906-3323, 2908-3323, 2913-3324, 2920-3320, 2920-3322, 2924-3318, 2934-3323, 2935-3320, 2939-2979, 2940-3320, 2945-3323, 2949-3313, 2951-3320, 2960-3318, 2960-3320, 2961-3320, 2967-3324, 2973-3314, 2973-3445, 2980-3323, 2987-3079, 2991-3225, 3003-3320, 3010-3320, 3012-3232, 3016-3307, 3016-3321, 3017-3320, 3018-3295, 3018-3323, 3026-3320, 3028-3443, 3030-3265, 3030-3315, 3030-3323, 3031-3281, 3040-3320, 3041-3323, 3051-3317, 3052-3272, 3053-3323, 3055-3320, 3059-3249, 3059-3320, 3142-3320, 3164-3251, 3164-3321, 3164-3323, 3275-3303, 3275-3325
64/1908201CBI/ 2929	1-305, 1-370, 1-373, 1-403, 1-404, 1-408, 1-429, 1-480, 1-511, 1-513, 1-521, 1-525, 1-529, 1-531, 1-537, 1-544, 1-551, 1-567, 1-580, 2-2929, 8-554, 11-502, 11-551, 18-371, 18-441, 18-447, 18-493, 23-339, 23-500, 27-492, 34-592, 95-718, 121-736, 136-737, 245-766, 264-659, 274-903, 394-675, 416-1063, 440-1055, 544-1002, 635-1263, 667-914, 684-1222, 719-1269, 742-1119, 771-1440, 888-1269, 986-1432, 1013-1269, 1065-1352, 1089-1271, 1119-1224, 1212-1432, 1245-1420, 1258-1650, 1338-1891, 1440-2243, 1447-1753, 1523-1955, 1600-2162, 1600-2204, 1625-1900, 1626-1900, 2014-2249, 2014-2253, 2014-2281, 2014-2421, 2075-2550, 2169-2619, 2339-2613, 2575-2929

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
65/2827615CB1/ 1923	1-567, 2-669, 3-390, 3-446, 3-480, 3-541, 4-115, 4-439, 4-529, 4-533, 4-580, 4-586, 4-602, 4-613, 8-624, 25-643, 71-554, 111-632, 112-677, 112-772, 148-484, 194-587, 194-627, 194-733, 195-634, 226-698, 231-859, 282-673, 282-840, 316-911, 324-929, 340-833, 341-832, 359-833, 368-911, 377-833, 378-911, 380-910, 384-825, 384-875, 384-932, 388-911, 390-899, 397-988, 412-911, 433-1012, 442-995, 448-833, 474-911, 496-988, 497-988, 498-911, 498-995, 498-1008, 498-1072, 508-1072, 542-988, 556-911, 556-1156, 556-1177, 557-995, 557-1014, 558-1156, 607-983, 616-1247, 653-913, 660-1156, 674-1247, 681-1240, 681-1247, 717-1330, 729-1247, 730-1239, 770-1012, 784-1324, 792-1247, 800-882, 800-888, 800-889, 800-902, 800-905, 800-921, 800-951, 800-974, 800-979, 800-987, 800-988, 800-1026, 800-1047, 800-1057, 800-1058, 800-1062, 800-1063, 800-1066, 800-1083, 800-1119, 800-1140, 800-1141, 800-1149, 800-1155, 800-1157, 800-1202, 800-1226, 800-1238, 800-1240, 802-897, 802-899, 802-988, 802-992, 802-1066, 802-1078, 802-1079, 802-1124, 802-1140, 807-1119, 817-967, 817-1046, 817-1155, 834-1311, 834-1324, 837-888, 837-1331, 840-888, 840-899, 840-902, 840-962, 840-973, 840-995, 840-996, 840-999, 840-1035, 840-1215, 840-1219, 840-1226, 840-1248, 840-1301, 840-1323, 850-878, 850-879, 850-883, 850-888, 850-902, 850-903, 850-910, 850-911, 850-928, 850-929, 850-962, 850-963, 850-981, 850-982, 850-986, 850-987, 850-988, 850-994, 850-995, 850-1011, 850-1042, 850-1046, 850-1048, 850-1057, 850-1065, 850-1066, 850-1124, 850-1130, 850-1132, 850-1141, 850-1149, 850-1155, 850-1163, 850-1181, 850-1194, 850-1212, 850-1214, 850-1219, 850-1226, 850-1240, 850-1244, 850-1246, 850-1278, 850-1287, 850-1298, 850-1308, 850-1310, 850-1322, 850-1324, 850-1335, 850-1379, 850-1394, 851-897, 852-1156, 852-1407, 857-1163, 860-1202, 861-901, 861-1058, 892-1119, 907-1130, 907-1240, 907-1371, 907-1506, 910-1490, 912-1309, 912-1332, 912-1382, 912-1406, 913-1408, 918-1371, 931-1371, 931-1412, 933-995, 933-1222, 933-1247, 933-1415, 934-966, 934-994, 934-995, 934-997, 934-1013, 934-1014, 934-1066, 934-1079, 934-1092, 934-1095, 934-1132, 934-1140, 934-1216, 934-1298, 934-1324, 934-1328, 934-1408, 934-1414, 938-1216, 946-1415, 948-1328, 952-1287, 968-1042, 968-1050, 968-1051, 968-1089, 968-1141, 968-1155, 968-1160, 968-1226, 968-1317, 968-1324, 968-1387, 968-1408, 968-1412, 968-1496, 972-1214, 974-1455, 975-1554, 976-1202, 985-1324, 989-1582,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	991-1328, 996-1393, 996-1416, 996-1466, 996-1469, 996-1490, 997-1490, 1002-1455, 1013-1569, 1017-1061, 1017-1245, 1017-1362, 1017-1476, 1017-1477, 1018-1050, 1018-1140, 1018-1155, 1018-1226, 1018-1250, 1018-1317, 1018-1412, 1018-1491, 1018-1496, 1020-1324, 1028-1490, 1032-1412, 1040-1250, 1054-1125, 1054-1151, 1054-1215, 1054-1240, 1054-1244, 1054-1250, 1054-1317, 1054-1393, 1054-1408, 1054-1490, 1054-1499, 1054-1575, 1056-1306, 1059-1698, 1060-1287, 1069-1408, 1071-1477, 1073-1666, 1080-1517, 1080-1521, 1080-1547, 1080-1550, 1080-1567, 1081-1575, 1086-1477, 1092-1151, 1092-1318, 1092-1384, 1092-1572, 1097-1464, 1102-1141, 1102-1379, 1102-1383, 1102-1550, 1102-1573, 1102-1575, 1112-1455, 1112-1575, 1116-1499, 1137-1548, 1147-1226, 1176-1202, 1177-1371, 1177-1623, 1181-1376, 1181-1490, 1181-1554, 1181-1584, 1181-1632, 1181-1742, 1186-1210, 1186-1218, 1186-1219, 1186-1224, 1186-1239, 1186-1249, 1186-1250, 1186-1324, 1186-1331, 1186-1344, 1186-1346, 1186-1401, 1186-1402, 1186-1408, 1186-1412, 1186-1470, 1186-1492, 1186-1510, 1186-1561, 1186-1575, 1186-1583, 1186-1586, 1186-1634, 1186-1644, 1186-1646, 1186-1657, 1186-1658, 1186-1659, 1186-1660, 1186-1661, 1188-1743, 1191-1660, 1193-1499, 1197-1412, 1231-1331, 1236-1753, 1236-1755, 1248-1583, 1248-1651, 1260-1667, 1268-1492, 1269-1292, 1269-1455, 1269-1477, 1269-1661, 1269-1707, 1269-1712, 1269-1729, 1269-1732, 1269-1743, 1270-1298, 1270-1301, 1270-1322, 1270-1330, 1270-1331, 1270-1333, 1270-1335, 1270-1408, 1270-1415, 1270-1418, 1270-1430, 1270-1467, 1270-1485, 1270-1500, 1270-1544, 1270-1548, 1270-1575, 1270-1645, 1270-1659, 1270-1660, 1270-1667, 1270-1674, 1270-1721, 1270-1740, 1270-1750, 1270-1751, 1270-1791, 1272-1544, 1272-1729, 1275-1743, 1277-1583, 1291-1908, 1304-1376, 1304-1386, 1304-1414, 1304-1415, 1304-1417, 1304-1425, 1304-1476, 1304-1477, 1304-1490, 1304-1496, 1304-1500, 1304-1569, 1304-1645, 1304-1655, 1304-1723, 1304-1730, 1308-1551, 1312-1477, 1312-1729, 1321-1660, 1325-1791, 1332-1751, 1338-1729, 1338-1791, 1338-1804, 1338-1923, 1344-1387, 1344-1392, 1344-1570, 1344-1636, 1344-1791, 1345-1380, 1345-1387, 1345-1415, 1345-1623, 1345-1714, 1351-1732, 1353-1574, 1353-1628, 1353-1791, 1354-1730, 1364-1791, 1366-1667, 1368-1730, 1382-1490, 1388-1470, 1388-1487, 1388-1490, 1388-1574, 1388-1575, 1388-1593, 1388-1646, 1388-1712, 1388-1729, 1388-1790, 1392-1634, 1398-1623, 1398-1655, 1398-1707, 1398-1791, 1399-1477, 1404-1791, 1405-1743, 1416-1732, 1416-1791, 1422-1790, 1422-1791, 1429-1668, 1435-1791, 1437-1713, 1437-1714,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1437-1791, 1438-1488, 1438-1593, 1438-1646, 1438-1718, 1438-1790, 1440-1729, 1440-1791, 1440-1825, 1443-1790, 1450-1730, 1450-1791, 1452-1790, 1472-1544, 1472-1555, 1472-1577, 1472-1585, 1472-1593, 1472-1644, 1472-1646, 1472-1660, 1472-1661, 1472-1668, 1472-1733, 1472-1791, 1473-1554, 1473-1751, 1474-1592, 1483-1575, 1513-1707, 1513-1791, 1519-1712, 1519-1790, 1519-1791, 1521-1729, 1522-1733, 1522-1790, 1522-1791, 1533-1791, 1534-1790, 1560-1730, 1564-1791, 1564-1825, 1582-1790, 1582-1791, 1584-1723, 1584-1729, 1584-1790, 1584-1791, 1592-1791, 1595-1658, 1596-1638, 1596-1644, 1596-1664, 1596-1668, 1596-1675, 1596-1729, 1596-1730, 1596-1743, 1596-1751, 1596-1763, 1596-1791, 1596-1792, 1597-1632, 1597-1659, 1597-1718, 1597-1719, 1597-1791, 1603-1790, 1603-1791, 1605-1791, 1606-1791, 1608-1791, 1611-1791, 1613-1791, 1616-1791, 1628-1723, 1628-1791, 1628-1791, 1640-1722, 1640-1728, 1640-1761, 1640-1789, 1640-1791, 1648-1753, 1656-1787, 1656-1791, 1677-1791, 1680-1791, 1681-1791, 1688-1791, 1689-1781, 1689-1791, 1690-1790, 1692-1791, 1697-1763, 1764-1790, 1765-1791
66/4304550CBI/ 5601	1-3689, 3032-3731, 3119-3798, 3120-3798, 3388-3798, 3432-3798, 3667-4262, 3679-4048, 3697-3989, 3699-4143, 4060-4469, 4060-4551, 4077-4335, 4079-4543, 4113-4598, 4129-4299, 4151-4424, 4158-4579, 4161-4579, 4163-4579, 4165-4587, 4171-4647, 4186-4517, 4211-4507, 4230-4514, 4235-4483, 4239-4615, 4325-4832, 4376-4888, 4418-4668, 4456-4728, 4456-4729, 4501-4702, 4528-4641, 4528-4759, 4536-5176, 4546-5156, 4641-5232, 4715-5293, 4746-5351, 4800-5277, 4809-5183, 4824-5241, 4841-5349, 4841-5510, 4851-5336, 4892-5411, 4895-5468, 4913-5429, 4916-5396, 4946-5196, 4946-5408, 4979-5398, 4995-5378, 5008-5262, 5008-5568, 5011-5275, 5019-5274, 5022-5271, 5023-5207, 5066-5601, 5136-5411
67/7473738CBI/ 2802	1-2802, 402-942, 839-1452, 855-1375, 1056-1553, 1070-1476, 1107-1700, 1247-1807, 1259-1708, 1307-1886, 1396-1511, 1397-2001, 1399-1657, 1399-1960
68/4447743CBI/ 2157	1-662, 1-713, 1-2157, 226-727, 498-1205, 507-973, 513-1349, 514-1318, 539-1246, 564-1090, 616-652, 616-673, 616-741, 616-764, 616-808, 616-931, 616-1009, 616-1013, 616-1023, 616-1091, 616-1093, 616-1094, 616-1098, 616-1099, 616-1800, 619-652, 619-673, 619-683, 619-723, 619-728, 619-755, 619-763, 619-801, 619-803, 619-806, 619-1015, 619-1181, 619-1182, 619-1394, 620-660, 620-668, 620-673, 620-754, 620-764, 620-781, 620-787, 620-791, 620-833, 620-839, 620-841, 620-847, 620-890, 620-896, 620-922, 620-925, 620-926, 620-929, 620-972, 620-975, 620-980, 620-1007, 620-1015, 620-1055, 620-1056, 620-1058, 620-1059, 620-1064, 620-1646, 620-1729, 620-1898, 621-838, 634-673, 634-806, 634-846, 634-847, 634-877, 634-1006, 634-1010, 634-1014, 634-1099, 634-1182, 634-1183, 636-847, 638-757, 638-890, 643-890, 661-937, 681-1392, 688-921, 688-1285, 688-1291, 693-847, 693-890, 701-812, 710-753, 710-758, 710-804, 710-807, 710-837, 710-846, 710-876, 710-890, 710-909, 710-931, 710-960, 710-976, 710-1099, 710-1178, 710-1259, 710-1308, 710-1310, 710-1311, 710-1464, 710-1982, 711-737, 711-740, 711-758, 711-781, 711-806, 711-812, 711-825, 711-841, 711-842, 711-845,

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
	711-847, 711-892, 711-923, 711-926, 711-931, 711-969, 711-971, 711-974, 711-1093, 711-1097, 711-1181, 711-1184, 711-1225, 712-764, 712-795, 712-842, 712-846, 712-847, 712-875, 712-891, 712-917, 712-961, 712-980, 712-1058, 712-1094, 712-1098, 712-1099, 712-1181, 712-1211, 712-1570, 714-846, 714-930, 714-1177, 714-1316, 715-811, 715-989, 715-1183, 715-1267, 715-1402, 715-1433, 715-2066, 719-763, 719-993, 719-1394, 724-974, 725-956, 730-1225, 732-1001, 732-1221, 733-896, 733-1010, 735-931, 735-974, 735-1058, 746-814, 746-974, 746-1015, 747-1064, 747-1225, 750-974, 766-1015, 771-1395, 774-890, 778-974, 778-1267, 783-1005, 783-1044, 783-1362, 783-1375, 784-838, 784-1073, 784-1350, 784-1514, 784-2066, 785-1258, 786-1015, 786-1261, 789-1045, 789-1195, 791-1015, 791-1349, 792-1099, 792-1400, 793-1060, 793-1434, 794-1015, 794-1174, 796-1285, 799-1310, 801-847, 803-1077, 803-1470, 808-1058, 809-1043, 816-1085, 816-1307, 822-1015, 822-1099, 830-1099, 831-1096, 831-1310, 851-1099, 865-922, 865-1161, 865-1454, 865-1498, 865-1570, 865-1601, 865-2066, 866-1342, 869-898, 869-974, 869-1058, 869-1089, 869-1127, 869-1430, 869-1434, 870-1099, 870-1345, 873-1094, 873-1184, 875-1099, 875-1433, 876-1394, 879-1058, 879-1099, 879-1142, 879-1518, 880-1225, 880-1349, 880-1350, 882-1434, 885-931, 887-1161, 887-1532, 892-1099, 893-1127, 900-1160, 900-1391, 903-1099, 903-1225, 913-982, 914-1221, 915-1225, 915-1394, 929-1229, 937-1099, 952-1426, 953-1006, 953-1058, 953-1173, 953-1180, 953-1435, 953-1517, 953-1519, 953-1562, 953-1685, 953-2065, 954-1219, 954-1310, 954-1429, 954-1462, 958-1181, 958-1352, 959-1215, 959-1218, 959-1511, 960-1478, 963-1098, 963-1183, 963-1644, 964-1310, 964-1433, 966-1568, 967-1519, 969-1015, 971-1181, 971-1601, 976-1225, 977-1183, 984-1253, 984-1473, 987-1225, 987-1229, 987-1310, 997-1066, 998-1305, 999-1316, 999-1478, 1002-1225, 1019-1267, 1021-1225, 1031-1094, 1036-1510, 1037-1090, 1037-1099, 1037-1257, 1037-1296, 1037-1352, 1037-1601, 1037-1603, 1037-1611, 1037-1647, 1037-1766, 1037-2066, 1038-1267, 1038-1513, 1041-1299, 1041-1625, 1043-1099, 1043-1267, 1048-1225, 1048-1268, 1048-1312, 1048-1394, 1048-1519, 1048-1532, 1048-1686, 1050-1652, 1051-1570, 1051-1601, 1055-1329, 1055-1688, 1060-1310, 1061-1295, 1063-1267, 1063-1601, 1068-1337, 1068-1559, 1071-1267, 1071-1310, 1071-1394, 1082-1389, 1083-1400, 1083-1570, 1114-1393, 1121-1225, 1121-1310, 1121-1341,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1121-1348, 1121-1682, 1121-1729, 1125-1383, 1125-1436, 1132-1178, 1132-1350, 1132-1396, 1132-1478, 1132-1601, 1132-1809, 1134-1597, 1134-1686, 1135-1350, 1135-1646, 1135-1685, 1139-1408, 1147-1350, 1147-1394, 1147-1776, 1152-1421, 1152-1601, 1166-1394, 1167-1478, 1167-1646, 1170-1394, 1171-1466, 1179-1715, 1189-1394, 1198-1815, 1202-1787, 1204-1570, 1206-1477, 1207-1563, 1207-1570, 1210-1570, 1212-1425, 1212-1432, 1212-1466, 1212-1763, 1212-1787, 1216-1350, 1216-1408, 1216-1433, 1216-1435, 1216-1466, 1216-1473, 1216-1594, 1216-1601, 1216-1709, 1216-1750, 1216-1822, 1216-1850, 1216-1896, 1216-2074, 1218-1681, 1218-1820, 1237-1310, 1237-1435, 1237-1478, 1237-1505, 1237-1725, 1237-1729, 1239-1477, 1239-1478, 1241-1394, 1242-1685, 1246-1433, 1246-1898, 1250-1478, 1250-1557, 1251-1568, 1251-1729, 1254-1478, 1271-1519, 1273-1478, 1278-1394, 1282-1787, 1282-1899, 1287-1509, 1287-1548, 1287-1873, 1287-1879, 1288-1399, 1288-1577, 1288-1867, 1288-1897, 1288-1989, 1288-2021, 1288-2066, 1289-1762, 1290-1519, 1290-1765, 1293-1549, 1293-1698, 1295-1519, 1295-1847, 1296-1646, 1296-1904, 1297-1562, 1297-1980, 1298-1435, 1298-1525, 1298-1678, 1298-1771, 1298-1906, 1298-1934, 1298-2066, 1300-1787, 1303-1814, 1305-1350, 1307-1581, 1307-1982, 1312-1563, 1313-1547, 1320-1589, 1320-1809, 1326-1519, 1326-1646, 1334-1601, 1335-1652, 1335-1814, 1357-1570, 1359-1593, 1360-1598, 1360-1957, 1360-1963, 1365-1478, 1365-1632, 1371-1983, 1372-1405, 1373-1601, 1374-1601, 1374-1849, 1376-1562, 1377-1633, 1377-1688, 1379-1601, 1379-1931, 1380-1729, 1380-1873, 1380-1988, 1381-1648, 1381-2064, 1384-1433, 1387-1898, 1389-1435, 1391-1665, 1391-2123, 1396-1646, 1397-1631, 1404-1673, 1404-1895, 1411-1646, 1418-1646, 1418-1725, 1422-1601, 1423-1723, 1431-1957, 1441-1646, 1453-1933, 1457-1814, 1458-1570, 1458-1601, 1458-1677, 1458-1684, 1458-1729, 1458-1733, 1458-1898, 1458-2041, 1458-2043, 1458-2066, 1459-1814, 1463-1719, 1463-1722, 1463-2117, 1464-2066, 1465-1687, 1465-2066, 1466-1661, 1466-1687, 1466-1846, 1466-1934, 1466-2065, 1466-2066, 1471-1519, 1473-2136, 1482-1729, 1482-1814, 1482-1982, 1483-1685, 1483-1687, 1483-1729, 1483-1868, 1483-2043, 1489-1757, 1489-1979, 1491-1729, 1491-1733, 1494-1959, 1502-1809, 1503-1820, 1503-1982, 1517-1771, 1537-1829, 1537-2018, 1537-2023, 1538-2014, 1541-1646, 1541-1729, 1541-1761, 1541-1800, 1541-2066, 1542-1771, 1542-2017, 1545-1801, 1545-1928, 1547-1771, 1547-2066, 1551-1687, 1551-1729, 1551-1772, 1551-1816, 1551-1930, 1551-2024, 1551-2050, 1551-2065, 1552-2098, 1554-2066, 1557-1605,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1559-1832, 1564-1814, 1564-2066, 1565-1799, 1572-1841, 1572-2063, 1573-1875, 1573-1898, 1573-2050, 1575-1771, 1575-1814, 1575-1898, 1585-1654, 1586-1893, 1587-1904, 1587-2157, 1590-1814, 1591-2087, 1607-1897, 1609-1814, 1624-2066, 1625-1678, 1625-1729, 1625-1845, 1625-1884, 1625-1937, 1625-2065, 1625-2066, 1626-1891, 1626-2066, 1629-1887, 1629-2022, 1631-1707, 1631-1880, 1636-1814, 1636-1852, 1636-1900, 1636-1982, 1636-2015, 1636-2062, 1636-2066, 1638-2061, 1639-2066, 1643-1917, 1643-2066, 1648-1898, 1649-1852, 1651-1890, 1651-2066, 1656-1925, 1656-2066, 1659-1897, 1659-1898, 1659-1982, 1670-1977, 1671-1988, 1671-2066, 1674-1898, 1675-1934, 1683-2074, 1693-1898, 1701-1814, 1702-2023, 1706-2065, 1708-2152, 1710-1981, 1711-2151, 1714-2141, 1716-1929, 1716-1934, 1716-1968, 1716-2065, 1716-2066, 1720-1934, 1720-1937, 1720-1969, 1720-1984, 1720-2066, 1720-2074, 1722-2023, 1722-2039, 1741-1967, 1741-1982, 1741-2009, 1741-2066, 1743-1934, 1743-1982, 1745-1898, 1746-2043, 1750-2001, 1750-2065, 1754-2061, 1755-2066, 1775-2065, 1777-1982, 1792-2066, 1793-1898, 1793-1903, 1793-2013, 1793-2020, 1793-2021, 1793-2066, 1794-2050, 1794-2065, 1797-2021, 1797-2041, 1799-2050, 1799-2065, 1800-2066, 1803-1982, 1803-1997, 1803-2023, 1803-2066, 1804-2066, 1806-2066, 1807-2023, 1811-2066, 1816-2069, 1817-2039, 1824-2066, 1827-2065, 1827-2066, 1827-2150, 1838-2066, 1838-2097, 1839-2066, 1842-2152, 1843-2066, 1860-2066, 1861-2023, 1861-2066, 1861-2150, 1868-1982, 1869-2066, 1869-2074, 1870-2066, 1875-2074, 1876-1930, 1876-2066, 1877-2066, 1878-2066, 1881-2024, 1881-2065, 1883-2066, 1883-2067, 1884-2066, 1885-2066, 1888-2050, 1891-2065, 1895-2074, 1900-2066, 1901-2065, 1903-2066, 1914-2023, 1914-2065, 1914-2066, 1919-2066, 1922-2043, 1922-2066, 1923-2065, 1926-2066, 1945-2066, 1950-2150, 1953-2066, 1954-2066, 1959-2066, 1960-2001, 1960-2014, 1961-2014, 1961-2066, 1962-2074, 1965-2021, 1967-2066, 1967-2074, 1971-2066, 1972-2021, 1974-2066, 1977-2023, 1979-2066, 1984-2066, 1985-2066, 1991-2066, 1992-2066, 1993-2066, 1995-2023, 1995-2061, 1995-2066, 2006-2066, 2011-2066, 2030-2066
69/7497554CB1/ 2104	1-728, 185-730, 296-2104, 495-877

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
70/7475843CB1/ 1451	1-336, 44-790, 45-354, 45-502, 45-549, 45-593, 45-641, 45-685, 45-734, 46-540, 57-700, 112-472, 112-578, 114-724, 117-240, 117-336, 117-1451, 118-376, 125-742, 125-769, 126-790, 127-394, 132-375, 133-398, 133-522, 133-792, 146-685, 154-707, 154-741, 173-702, 241-1451, 337-1451, 338-765, 365-800, 365-896, 372-981, 383-1151, 399-1129, 471-1093, 563-1264, 565-948, 582-741, 582-826, 596-844, 596-903, 596-994, 596-1078, 637-1442, 685-826, 685-897, 685-899, 685-994, 685-995, 685-1045, 685-1067, 685-1155, 685-1162, 685-1340, 689-924, 689-961, 689-1071, 689-1245, 692-1097, 696-1245, 697-924, 701-1245, 712-1129, 754-1450, 763-983, 763-1245, 769-980, 769-995, 769-1177, 769-1210, 769-1231, 769-1424, 773-1045, 773-1340, 776-1340, 794-1451, 857-1064, 857-1065, 857-1114, 857-1181, 857-1210, 857-1239, 857-1245, 857-1295, 857-1319, 857-1329, 857-1340, 857-1413, 864-1239, 915-1451, 931-1149, 931-1295, 931-1413, 932-1232, 932-1414, 937-1148, 937-1198, 937-1345, 937-1379, 937-1403, 937-1414, 941-1424, 944-1424, 945-1347, 969-1282, 1008-1414, 1016-1161, 1016-1210, 1016-1316, 1016-1414, 1016-1424, 1025-1233, 1025-1372, 1025-1413, 1028-1413, 1032-1414, 1047-1317, 1047-1414, 1084-1413, 1085-1245, 1085-1330, 1086-1149, 1086-1150, 1086-1153, 1086-1245, 1086-1330, 1099-1414, 1100-1153, 1100-1330, 1105-1161, 1105-1344, 1105-1414, 1109-1414, 1113-1330, 1116-1414, 1131-1344, 1131-1414, 1132-1414, 1181-1414, 1183-1414, 1193-1414, 1218-1414, 1252-1414, 1255-1414, 1273-1368, 1273-1414, 1288-1330, 1289-1414, 1357-1414, 1361-1414, 1364-1414
71/6319550CB1/ 1609	1-464, 1-1038, 69-480, 134-490, 147-1038, 491-684, 499-942, 501-963, 501-964, 502-977, 511-865, 516-795, 516-810, 516-819, 569-855, 695-936, 695-1238, 851-1370, 881-1393, 964-1190, 1013-1609, 1047-1609, 1086-1408, 1103-1393, 1106-1530, 1185-1609, 1246-1507, 1295-1525, 1313-1591
72/7510064CB1/ 2840	1-286, 1-2816, 106-638, 111-616, 116-333, 121-344, 121-469, 173-602, 173-677, 173-735, 180-501, 226-597, 228-597, 371-808, 372-657, 394-873, 1183-1567, 1183-1569, 1377-1916, 1777-2047, 2152-2382, 2152-2437, 2179-2828, 2180-2766, 2186-2410, 2186-2430, 2186-2513, 2197-2838, 2214-2517, 2235-2840, 2261-2828, 2267-2442, 2289-2579, 2310-2563, 2335-2791, 2337-2760, 2345-2568, 2348-2616, 2350-2820, 2353-2635, 2356-2816, 2360-2818, 2362-2642, 2366-2816, 2368-2817, 2369-2778, 2382-2816, 2386-2817, 2392-2774, 2396-2827, 2399-2766, 2403-2662, 2404-2601, 2404-2823, 2433-2835, 2443-2694, 2443-2774, 2443-2814, 2471-2816, 2477-2774, 2499-2811, 2507-2811, 2509-2816, 2515-2816, 2522-2816, 2540-2792, 2540-2834, 2547-2816, 2604-2798, 2609-2770, 2609-2811, 2610-2812, 2610-2828, 2618-2838, 2680-2838

PF-1031 PCT

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
38	7490301CB1	LUNGNOT25
39	2383223CB1	COLNDIY01
40	3495982CB1	BRAIFER05
41	7477891CB1	CONNTMA01
42	72688352CB1	KIDNNOT05
44	7489744CB1	HEARFEF03
45	3363382CB1	BRAUNOR01
46	7491148CB1	SINIDME01
47	8126343CB1	PANCTUT01
48	7044055CB1	SINTNOR01
49	7493424CB1	BMARTXE01
50	1482140CB1	PROSNOT14
51	394992CB1	MIXDDIE02
52	5093550CB1	BRAUTDR03
54	1706514CB1	UTRSTMR02
55	7488247CB1	THP1AZT01
56	1427269CB1	LUNGFET04
57	103135CB1	THP1NOT03
58	1907346CB1	PLACFEB01
59	3041036CB1	SINTNOT19
60	3856879CB1	UTRETMCO1
61	4178665CB1	BRAINOT22
62	7493326CB1	TYMNOT02
63	1553836CB1	293TF2T01
64	1908201CB1	SKINDIA01
65	2827615CB1	PLACFER06
66	4304550CB1	COLATMT01
67	7473738CB1	LUNGTUT13
68	4447743CB1	BRAINOT12
69	7497554CB1	GBLADIE01
70	7475843CB1	THP1UNT03
71	6319550CB1	BRADDIR01
72	7510064CB1	STOMNOT01

Table 6

Library	Vector	Library Description
293TF2T01	pINCY	Library was constructed using RNA isolated from a treated, transformed embryonic cell line (293-EBNA) derived from kidney epithelial tissue. The cells were treated with 5-aza-2'-deoxycytidine and transformed with adenovirus 5 DNA.
BMARTXE01	pINCY	This 5' biased random primed library was constructed using RNA isolated from treated SH-SY5Y cells derived from a metastatic bone marrow neuroblastoma, removed from a 4-year-old Caucasian female (Scherling AG). The medium was MEM/HAM'S F12 with 10% fetal calf serum. After reaching about 80% confluency cells were treated with 6-Hydroxydopamine (6-OHDA) at 100 microM for 8 hours.
BRADDIR01	pINCY	Library was constructed using RNA isolated from diseased choroid plexus tissue of the lateral ventricle, removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident.
BRAIFER05	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRAINOT12	pINCY	Library was constructed using RNA isolated from brain tissue removed from the right frontal lobe of a 5-year-old Caucasian male during a hemispherectomy. Pathology indicated extensive polymicrogyria and mild to moderate gliosis (predominantly subpial and subcortical), which are consistent with chronic seizure disorder. Family history included a cervical neoplasm.
BRAINOT22	pINCY	Library was constructed using RNA isolated from right temporal lobe tissue removed from a 45-year-old Black male during a brain lobectomy. Pathology for the associated tumor tissue indicated dysembryoplastic neuroepithelial tumor of the right temporal lobe. The right temporal region dura was consistent with calcifying pseudotumor of the neuraxis. Family history included obesity, benign hypertension, cirrhosis of the liver, obesity, hyperlipidemia, cerebrovascular disease, and type II diabetes.
BRAUNOR01	pINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by

Table 6

Library	Vector	Library Description
BRAUTDR03	PCDNA2.1	<p>reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.</p> <p>This random primed library was constructed using RNA isolated from pooled globus pallidus and substantia innominata tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.</p>
COLATMT01	pINCY	<p>Library was constructed using RNA isolated from ascending and transverse colon tissue removed from a 55-year-old Caucasian male during right hemicolectomy, incidental appendectomy, and permanent colostomy. Pathology for the matched tumor tissue indicated invasive grade 3 adenocarcinoma that formed a circumferential mass in the ascending colon. Patient history included benign hypertension, anxiety, abnormal blood chemistry, blepharitis, heart block, osteoporosis, acne, and hyperplasia of prostate. Previous surgeries included adenotonsillectomy. Patient medications included Ativan, multivitamins, and minerals. Family history included prostate cancer in the father; and acute myocardial infarction, stroke, and atherosclerotic coronary artery disease in the grandparent(s).</p>
COLNDIY01	pINCY	<p>This large size-fractionated and normalized library was constructed using pooled cDNA from two different donors. cDNA was generated using RNA isolated from normal colon tissue removed from a 16-year-old black male (donor A) who died in a motor vehicle accident and diseased ascending colon polyp removed from a 26-year-old Caucasian male (donor B) during a partial colectomy, permanent colostomy, and incidental appendectomy. Pathology for donor B indicated moderately to severely active Crohn disease, involving the terminal ileum, cecum, and ascending colon. Grossly, the specimens showed patchy involvement, creeping fat, stricture, and numerous inflammatory pseudopolyps. Microscopically, the specimen showed transmural inflammation with skip areas, mural fibrosis, fissuring ulceration, and lymphoid aggregates present in all layers of the bowel wall. Medical history included positive serology for cytomegalovirus in donor A; and regional enteritis of the small intestine, blood in stool, functional diarrhea, abdominal pain, fevers, chills, and anemia in donor B. Family history included regional enteritis of the small intestine and enteritis of the small intestine in donor B. 0.6 million independent clones from this size-selected library were normalized in one round using conditions adapted from Soares et al. (1994; Proc. Natl. Acad. Sci. USA 91:9228-9232) and Bonaldo et al. (1996; Genome Research 6:791).</p>

Table 6

Library	Vector	Library Description
CONNTMA01	PSPORT1	This amplified library was constructed using RNA made from pooled cDNA from two donors. cDNA was generated using mRNA isolated from supraglottic soft tissue removed from a 61-year-old Caucasian male (donor A) during complete laryngectomy, radical neck dissection, and cricopharyngeal myotomy; and from retroperitoneal soft tissue removed from a 17-year-old Caucasian female (donor B) during excision of lesion of soft tissue. Pathology indicated negative margin for A and B. Pathology for the associated tumor tissue (A) indicated invasive grade 3 squamous cell carcinoma involving the epiglottis centrally. Pathology for the associated tumor tissue (B) indicated hemangiopericytoma in the retroperitoneal region, the spine (L-3-4-5), and the L-3 left nerve root. The patient presented with dysphagia (A), malignant neoplasm of the soft tissue and lumbar nerve, anxiety state and joint pain (B). Patient history included osteoporosis and acute sinusitis (A) and tobacco abuse (B). Previous surgeries included vocal cordectomy, indirect inguinal hernia repair, and vasectomy (A) and adenotonsillectomy (B). Patient medications included aspirin and Digoxin (A) and Ortho-Sept (B). Family history included depressive disorder in the sibling(s) and type II diabetes in the grandparent(s) (A); and alcohol abuse, benign hypertension, acute myocardial infarction, and atherosclerotic coronary artery disease in the father; and alcohol abuse and benign hypertension in the grandparent(s) (B).
GBLADIE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from diseased gallbladder tissue removed from a 55-year-old Caucasian female during laparoscopic cholecystectomy. Pathology indicated chronic cholecystitis and cholelithiasis (greater than 100 stones). The patient presented with cholelithiasis, abdominal pain, and tremors. Patient history included benign hypertension, Morton's neuroma, facial hirsutism, normal delivery, and tobacco abuse in remission. Previous surgeries included total abdominal hysterectomy, bilateral salpingo-oophorectomy, and adenotonsillectomy. Patient medications included Inderal and Premarin. Family history included breast cancer and ALS in the mother; chronic leukemia and ARDS in the father; breast cancer in the sibling(s); and atherosclerotic coronary artery disease in the grandparent(s).
HEARFE03	PCMV-ICIS	Library was constructed using RNA isolated from heart tissue removed from a Caucasian male fetus who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation. Serologies were negative.
KIDNNOT05	PSPORT1	Library was constructed using RNA isolated from the kidney tissue of a 2-day-old Hispanic female, who died from cerebral anoxia. Family history included congenital heart disease.
LUNGFET04	pINCY	Library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 17 weeks' gestation from anencephalus.
LUNGNOT25	pINCY	Library was constructed using RNA isolated from lung tissue removed from a 33-year-old Caucasian male. Pathology for the associated tumor tissue indicated endobronchial carcinoid tumor. Patient history included deficiency anemia, Dengue fever, pneumonia, basal cell skin cancer, and presumed TB through cultures. Family history included leukemia, cerebrovascular disease, atherosclerotic coronary artery disease, and hyperlipidemia.

Table 6

Library	Vector	Library Description
LUNGTUT13	pINCY	Library was constructed using RNA isolated from tumorous lung tissue removed from the right upper lobe of a 47-year-old Caucasian male during a segmental lung resection. Pathology indicated invasive grade 3 (of 4) adenocarcinoma. Family history included atherosclerotic coronary artery disease, and type II diabetes.
MIXDDIE02	PBK-CMV	This 5' biased random primed library was constructed using pooled cDNA from seven donors. cDNA was generated using mRNA isolated from brain tissue removed from two Caucasian male fetuses who died after 23 weeks gestation from hypoplastic left heart (A) and prematurity (B); from posterior hippocampus from a 55-year-old male who died from COPD (C); from cerebellum, corpus callosum, thalamus and temporal lobe tissue from a 57-year-old Caucasian male who died from a CVA (D); from dentate nucleus and vermis from an 82-year-old Caucasian male who died from a myocardial infarction (E); from pituitary gland from a 74-year-old Caucasian female who died from a myocardial infarction (F) and vermis tissue from a 77-year-old Caucasian female who died from pneumonia (G). For donor C, pathology indicated mild lateral ventricular enlargement. For donor F, pathology indicated moderate Alzheimer's disease, recent multiple infarctions involving left thalamus, left parietal and occipital lobes (microscopic) and right cerebellum (gross), mild atherosclerosis involving middle cerebral arteries bilaterally and mild cerebral amyloid angiopathy. For donor G, pathology indicated severe Alzheimer's disease, mild atherosclerosis involving the middle cerebral and basilar arteries, and cerebral atrophy consistent with Alzheimer's disease. For donor D, patient history included Huntington's chorea. Donor E was taking nitroglycerin and dopamine; donor F was taking Lopressor, heparin, ceftriaxone, captopril, Isordil, nitroglycerin, Clinoril, Ecotrin and tacrine; and donor G was taking insulin.
PANCTUT01	pINCY	Library was constructed using RNA isolated from pancreatic tumor tissue removed from a 65-year-old Caucasian female during radical subtotal pancreatectomy. Pathology indicated an invasive grade 2 adenocarcinoma. Patient history included type II diabetes, osteoarthritis, cardiovascular disease, benign neoplasm in the large bowel, and a cataract. Previous surgeries included a total splenectomy, cholecystectomy, and abdominal hysterectomy. Family history included cardiovascular disease, type II diabetes, and stomach cancer.
PLACFEB01	pINCY	Library was constructed using pooled cDNA from two different donors. cDNA was generated using RNA isolated from placenta tissue removed from a Caucasian fetus (donor A), who died after 16 weeks' gestation from fetal demise and hydrocephalus; and a Caucasian male fetus (donor B), who died after 18 weeks' gestation from fetal demise. Patient history included umbilical cord wrapped around the head (3 times) and the shoulders (1 time) in donor A. Serology was positive for anti-CMV in donor A. Family history included multiple pregnancies and live births, and an abortion in donor A.
PLACFER06	pINCY	This random primed library was constructed using RNA isolated from placental tissue removed from a Caucasian fetus who died after 16 weeks' gestation from fetal demise and hydrocephalus. Patient history included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV. Family history included multiple pregnancies and live births, and an abortion.

Table 6

Library	Vector	Library Description
PROSNOT14	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 60-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). Patient history included a kidney cyst and hematuria. Family history included benign hypertension, cerebrovascular disease, and arteriosclerotic coronary artery disease.
SINIDME01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from diseased ileum tissue removed from a 29-year-old Caucasian female during jejunostomy. Pathology indicated mild chronic inflammation. The patient presented with ulcerative colitis. Patient history included a benign neoplasm of the large bowel. Patient medications included Asacol, Rowasa, Clomid and Pergonal. Family history included benign hypertension in the mother, and colon cancer and cerebrovascular accident in the grandparent(s).
SINTNOR01	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 31-year-old Caucasian female during Roux-en-Y gastric bypass. Patient history included clinical obesity.
SINTNOT19	pINCY	Library was constructed using RNA isolated from small intestine tissue removed from a 8-year-old Black male, who died from anoxia. Serologies were negative. Patient medications included DDVP, Versed, and labetalol.
SKINDIA01	PSPORT1	This amplified library was constructed using RNA isolated from diseased skin tissue removed from 1 female and 4 males during skin biopsies. Pathologies indicated tuberculoid and lepromatous leprosy.
STOMNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the stomach tissue of a 55-year-old Caucasian male, who died from cardiopulmonary arrest.
THP1AZT01	pINCY	Library was constructed using RNA isolated from THP-1 promonocyte cells treated for three days with 0.8 micromolar 5-aza-2'-deoxycytidine. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (Int. J. Cancer (1980) 26:171).
THPINOT03	pINCY	Library was constructed using RNA isolated from untreated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (Int. J. Cancer (1980) 26:171).
THPIUNT03	pINCY	Library was constructed using RNA isolated from untreated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (Int. J. Cancer (1980) 26:171). The library was linearized and recircularized to select for insert containing clones.
TLYMNOT02	PBLUESCRIPT	Library was constructed using RNA isolated from non-adherent peripheral blood mononuclear cells. The blood was obtained from unrelated male and female donors and treated with LPS for 0 hours.

Table 6

Library	Vector	Library Description
UTRETMCO1	pINCY	This large size-fractionated library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from endometrial tissue removed from a 32-year-old Caucasian female (donor A) during total abdominal hysterectomy, bilateral salpingo-oophorectomy, and cystocele repair; and from endometrial tissue removed from a 48-year-old Caucasian female (donor B) during a vaginal hysterectomy, rectocele repair, and bilateral salpingo-oophorectomy. Pathology for donor A indicated the endometrium was in the proliferative phase. The right ovary showed a corpus luteal cyst. For donor B, pathology indicated chronic cervicitis and the endometrium was weakly proliferative. The right ovary and specimen from the peritoneum indicated endometriosis focally involving the surface of the right ovary and the peritoneum. Pathology for the matched tumor tissue indicated a single submucosal leiomyoma, which exhibited extensive hyalin change with hyalin-type necrosis. The left ovary contained a corpus luteum cyst. Donor A presented with abdominal pain, stress incontinence, and dysmenorrhea. Patient history included hemorrhagic ovarian cysts, uterine endometriosis, normal delivery, and cesarean deliveries. Donor B presented with metrorrhagia, extrinsic asthma, depressive disorder, and anxiety state. Patient history included alcohol abuse, hyperlipidemia, a normal delivery, tobacco abuse in remission, and meningitis. Patient medications (B) included Prozac, Trazodone, Clorazepate, and Medrol.
UTRSTMR02	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from endometrial tissue removed from a 32-year-old female (donor A) and using mRNA isolated from myometrium removed from a 45-year-old female (donor B) during vaginal hysterectomy and bilateral salpingo-oophorectomy. In donor A, pathology indicated the endometrium was secretory phase. The cervix showed severe dysplasia (CIN III) focally involving the squamocolumnar junction at the 1, 6 and 7 o'clock positions. Mild koilocytotic dysplasia was also identified within the cervix. In donor B, pathology for the matched tumor tissue indicated multiple (23) subserosal, intramural, and submucosal leiomyomata. Patient history included stress incontinence, extrinsic asthma without status asthmaticus and normal delivery in donor B. Family history included cerebrovascular disease, depression, and atherosclerotic coronary artery disease in donor B.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value=1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.0E-6 Assembled ESTs: fasta Identity=95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART, and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART, or TIGRFAM hits: Probability value=1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score > GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length=56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
72	7510064	1297845H1	SNP00052114	20	2205	G	G	A	D729	0.96	0.9647816	0.9896677	n/a
72	7510064	1482140H1	SNP00052114	208	2205	G	G	A	D729	0.96	0.9647816	0.9896677	n/a
72	7510064	1563212H1	SNP00052114	90	2205	G	G	A	D729	0.96	0.9647816	0.9896677	n/a
72	7510064	1688372H1	SNP00052114	157	2205	G	G	A	D729	0.96	0.9647816	0.9896677	n/a
72	7510064	1692412H1	SNP00052114	20	2205	G	G	A	D729	0.96	0.9647816	0.9896677	n/a
72	7510064	2114157H1	SNP00052114	202	2205	G	G	A	D729	0.96	0.9647816	0.9896677	n/a
72	7510064	2358472H1	SNP00052114	145	2205	G	G	A	D729	0.96	0.9647816	0.9896677	n/a
72	7510064	2404442H1	SNP00027050	53	1281	T	C	T	S421	n/a	n/a	n/a	n/a
72	7510064	2890337H1	SNP00027050	8	1281	C	C	T	P421	n/a	n/a	n/a	n/a
72	7510064	2952667H1	SNP00052114	175	2205	G	G	A	D729	0.96	0.9647816	0.9896677	n/a
72	7510064	2956748H1	SNP00027050	61	1280	C	C	T	N420	n/a	n/a	n/a	n/a
72	7510064	2991071H1	SNP00073508	174	174	C	C	T	R52	n/a	n/a	n/a	n/a
72	7510064	3120215H1	SNP00052114	114	2205	G	G	A	D729	0.96	0.9647816	0.9896677	n/a
72	7510064	3436710H1	SNP00052114	55	2202	G	G	A	V728	0.96	0.9647816	0.9896677	n/a
72	7510064	4165273H1	SNP00052114	62	2204	G	G	A	L728	0.96	0.9647816	0.9896677	n/a
72	7510064	4165348H1	SNP00052114	61	2203	G	G	A	R728	0.96	0.9647816	0.9896677	n/a
72	7510064	4825752H1	SNP00052114	226	2200	G	G	A	S727	0.96	0.9647816	0.9896677	n/a
72	7510064	5756162H1	SNP00027050	206	1281	C	C	T	P421	n/a	n/a	n/a	n/a
72	7510064	5800564H1	SNP00073508	69	174	T	C	T	stop52	n/a	n/a	n/a	n/a
72	7510064	6118889H1	SNP00052114	180	2205	G	G	A	D729	0.96	0.9647816	0.9896677	n/a
72	7510064	6118916H1	SNP00052114	180	2205	G	G	A	D729	0.96	0.9647816	0.9896677	n/a
72	7510064	6128071H1	SNP00052114	181	2205	G	G	A	D729	0.96	0.9647816	0.9896677	n/a
72	7510064	6601515H1	SNP00052114	306	2205	A	G	A	N729	0.96	0.9647816	0.9896677	n/a
72	7510064	6910636J1	SNP00073510	419	1521	G	G	A	E501	n/d	n/d	n/d	n/d
72	7510064	6975819H1	SNP00073509	185	757	C	C	T	A246	0.86	0.8019807	0.8498107	0.8613485
72	7510064	6975819H1	SNP00149328	277	665	G	G	A	T215	n/a	n/a	n/a	n/a
72	7510064	7200316H1	SNP00073510	416	1521	G	G	A	E501	n/d	n/d	n/d	n/d
72	7510064	7453284H1	SNP00073510	161	1521	G	G	A	E501	n/d	n/d	n/d	n/d

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
72	7510064	7465908H1	SNP00073510	504	1521	G	G	A	E501	n/d	n/d	n/d	n/d
72	7510064	7466772H1	SNP00073508	64	174	C	C	T	R52	n/a	n/a	n/a	n/a
72	7510064	7714002J1	SNP00073509	533	757	T	C	T	V246	0.86	0.8019807	0.8498107	0.8613485
72	7510064	7715037J1	SNP00073510	330	1521	G	G	A	E501	n/d	n/d	n/d	n/d
72	7510064	7715660J1	SNP00073510	244	1521	G	G	A	E501	n/d	n/d	n/d	n/d
72	7510064	7753158H1	SNP00073509	172	757	C	C	T	A246	0.86	0.8019807	0.8498107	0.8613485
72	7510064	7753158H1	SNP00149328	80	665	G	G	A	T215	n/a	n/a	n/a	n/a

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

- 5 a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36,
- b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3-7, SEQ ID NO:10-21, SEQ ID NO:24-30, SEQ ID NO:32, SEQ ID NO:34, and SEQ ID NO:36,
- 10 c) a polypeptide comprising a naturally occurring amino acid sequence at least 91% identical to the amino acid sequence of SEQ ID NO:35,
- d) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO:9,
- e) a polypeptide comprising a naturally occurring amino acid sequence at least 97% identical to the amino acid sequence of SEQ ID NO:22,
- 15 f) a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2 and SEQ ID NO:33,
- g) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:23 and SEQ ID NO:31,
- 20 h) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, and
- i) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-36.
- 25

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36.

30 3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from

the group consisting of SEQ ID NO:37-72.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

5

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

10

9. A method of producing a polypeptide of claim 1, the method comprising:

- a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
- b) recovering the polypeptide so expressed.

15

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-36.

20

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-72,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:37-43 and SEQ ID NO:46-71,
- c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 92% identical to the polynucleotide sequence of SEQ ID NO:72,
- d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 95% identical to the polynucleotide sequence of SEQ ID NO:45,
- e) a polynucleotide complementary to a polynucleotide of a),
- f) a polynucleotide complementary to a polynucleotide of b),
- g) a polynucleotide complementary to a polynucleotide of c),

25

30

- h) a polynucleotide complementary to a polynucleotide of d), and
- i) an RNA equivalent of a)-h).

5 13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- 10 a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
- 15 b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

20 16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

25

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

30 18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-36.

19. A method for treating a disease or condition associated with decreased expression of functional NAAP, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

5

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

10 22. A method for treating a disease or condition associated with decreased expression of functional NAAP, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- 15 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

20

25. A method for treating a disease or condition associated with overexpression of functional NAAP, comprising administering to a patient in need of such treatment a composition of claim 24.

25 26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

30

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of NAAP in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide

complex, and

- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

- a) a chimeric antibody,  
b) a single chain antibody,  
c) a Fab fragment,  
d) a F(ab')<sub>2</sub> fragment, or  
e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of NAAP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of NAAP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, or an immunogenic fragment thereof, under conditions to elicit an antibody response,  
b) isolating antibodies from said animal, and  
c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-36, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific

binding of the antibody and the polypeptide, and

- b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-36.

5

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

10

- a) labeling the polynucleotides of the sample,  
b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and  
c) quantifying the expression of the polynucleotides in the sample.

15

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

20

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

25

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

30

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

5 71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

10

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

15

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

20

79. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.

80. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.

25

81. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

82. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:27.

83. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:28.

30

84. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:29.

85. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:30.

86. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:31.

87. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:32.

5 88. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:33.

89. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:34.

90. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:35.

10

91. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

15

93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

20

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

25

98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

100. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:45.

30

101. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:46.

102. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:47.

103. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:48.

104. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:49.

5 105. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:50.

106. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:51.

107. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:52.

10

108. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:53.

109. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:54.

15

110. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:55.

111. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:56.

112. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:57.

20

113. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:58.

114. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:59.

25

115. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:60.

116. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:61.

117. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:62.

30

118. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:63.

119. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:64.

120. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:65.

121. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:66.

5 122. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:67.

123. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:68.

124. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:69.

10

125. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:70.

126. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:71.

15

127. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:72.

<110> INCYTE GENOMICS, INC.

GANDHI, Ameena R.  
SWARNAKER, Anita  
HAFALIA, April J.A.  
WARREN, Bridget A.  
EMERLING, Brooke M.  
ARVIZU, Chandra S.  
ISON, Craig H.  
HONCHELL, Cynthia D.  
LEE, Ernestine A.  
YUE, Henry  
FORSYTHE, Ian J.  
RAMKUMAR, Jayalaxmi  
GRIFFIN, Jennifer A.  
YANG, Junming  
SANJANWALA, Madhu M.  
BAUGHN, Mariah R.  
BOROWSKY, Mark L.  
YAO, Monique G.  
WALIA, Narinder K.  
BANDMAN, Olga  
LAL, PREETI G.  
BECHA, Shanya D.  
LEE, Soo Yeun  
RICHARDSON, Thomas W.  
ELLIOTT, Vicki S.  
LUO, Wen  
TANG, Y. Tom  
ZEBARJADIAN, Yeganeh  
LU, Yan

<120> Nucleic Acid-Associated Proteins

<130> PF-1031 PCT

<140> To Be Assigned

<141> Herewith

<150> US 60/300,518

<151> 2001-06-22

<150> US 60/301,787

<151> 2001-06-29

<150> US 60/301,792

<151> 2001-06-29

<150> US 60/301,892

<151> 2001-06-29

<150> US 60/301,893

<151> 2001-06-29

<150> US 60/303,405

<151> 2001-07-06

<150> US 60/303,442

<151> 2001-07-06

<150> US 60/364,438

<151> 2002-03-12

<160> 72

&lt;150&gt; US 60/303,442

&lt;151&gt; 2001-07-06

&lt;150&gt; US 60/364,438

&lt;151&gt; 2002-03-12

&lt;160&gt; 72

&lt;170&gt; PERL Program

&lt;210&gt; 1

&lt;211&gt; 304

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7490148CD1

&lt;400&gt; 1

Met	Ser	Arg	Ser	Phe	Tyr	Val	Asp	Ser	Leu	Ile	Ile	Lys	Asp	Thr
1				5					10					15
Ser	Arg	Pro	Ala	Pro	Ser	Leu	Pro	Glu	Pro	His	Pro	Gly	Pro	Asp
				20					25					30
Phe	Phe	Ile	Pro	Leu	Gly	Met	Pro	Pro	Pro	Leu	Val	Met	Ser	Val
				35					40					45
Ser	Gly	Pro	Gly	Cys	Pro	Ser	Arg	Lys	Ser	Gly	Ala	Phe	Cys	Val
				50					55					60
Cys	Pro	Leu	Cys	Val	Thr	Ser	His	Leu	His	Ser	Ser	Arg	Gly	Ser
				65					70					75
Val	Gly	Pro	Ala	Ser	Gly	Gly	Ala	Gly	Pro	Gly	Phe	Pro	Gly	Pro
				80					85					90
Gly	Asp	Ser	Gly	Val	Ala	Gly	Pro	Ala	Gly	Ala	Leu	Pro	Leu	Leu
				95					100					105
Lys	Gly	Gln	Phe	Ser	Ser	Ala	Pro	Gly	Asp	Ala	Gln	Phe	Cys	Pro
				110					115					120
Arg	Val	Asn	His	Ala	His	His	His	His	His	Pro	Pro	Gln	His	His
				125					130					135
His	His	His	His	Gln	Pro	Gln	Gln	Pro	Gly	Ser	Ala	Ala	Ala	Ala
				140					145					150
Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Leu	Gly	His
				155					160					165
Pro	Gln	His	His	Ala	Pro	Val	Cys	Thr	Ala	Thr	Thr	Tyr	Asn	Val
				170					175					180
Ala	Asp	Pro	Arg	Arg	Phe	His	Cys	Leu	Thr	Met	Gly	Gly	Ser	Asp
				185					190					195
Ala	Ser	Gln	Val	Pro	Asn	Gly	Lys	Arg	Met	Arg	Thr	Ala	Phe	Thr
				200					205					210
Ser	Thr	Gln	Leu	Leu	Glu	Leu	Glu	Arg	Glu	Phe	Ser	Ser	Asn	Met
				215					220					225
Tyr	Leu	Ser	Arg	Leu	Arg	Arg	Ile	Glu	Ile	Ala	Thr	Tyr	Leu	Asn
				230					235					240
Leu	Ser	Glu	Lys	Gln	Val	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Val
				245					250					255
Lys	His	Lys	Lys	Glu	Gly	Lys	Gly	Thr	Gln	Arg	Asn	Ser	His	Ala
				260					265					270
Gly	Cys	Lys	Cys	Val	Gly	Ser	Gln	Val	His	Tyr	Ala	Arg	Ser	Glu

	275	280	285
Asp Glu Asp Ser Leu Ser Pro Ala Ser Ala Asn Asp Asp Lys Glu			
	290	295	300
Ile Ser Pro Leu			

&lt;210&gt; 2

&lt;211&gt; 198

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7490301CD1

&lt;400&gt; 2

Met Glu Thr Gly Arg Gln Ala Gly Val Ser Ala Glu Met Phe Ala		
1 5 10 15		
Met Pro Arg Asp Leu Lys Gly Ser Asn Lys Asp Gly Ile Pro Glu		
20 25 30		
Asp Leu Asp Gly Asn Leu Glu Glu Pro Arg Asp Gln Glu Gly Glu		
35 40 45		
Leu Arg Ser Glu Asp Val Met Asp Leu Thr Glu Gly Asp Asn Glu		
50 55 60		
Ala Ser Ala Ser Ala Pro Pro Ala Ala Lys Arg Arg Lys Thr Asp		
65 70 75		
Thr Lys Gly Lys Lys Glu Arg Lys Pro Thr Val Asp Ala Glu Glu		
80 85 90		
Ala Gln Arg Met Thr Thr Leu Leu Ser Ala Met Ser Glu Glu Gln		
95 100 105		
Leu Ser Arg Tyr Glu Val Cys Arg Arg Ser Ala Phe Pro Lys Ala		
110 115 120		
Cys Ile Ala Gly Leu Met Arg Ser Ile Thr Gly Arg Ser Val Ser		
125 130 135		
Glu Asn Val Ala Ile Ala Met Ala Gly Ile Ala Lys Val Phe Val		
140 145 150		
Gly Glu Val Val Glu Glu Ala Leu Asp Val Cys Glu Met Trp Gly		
155 160 165		
Glu Met Pro Pro Leu Gln Pro Lys His Leu Arg Glu Ala Val Arg		
170 175 180		
Arg Leu Lys Pro Lys Gly Leu Phe Pro Asn Ser Asn Tyr Lys Lys		
185 190 195		
Ile Met Phe		

&lt;210&gt; 3

&lt;211&gt; 576

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2383223CD1

&lt;400&gt; 3

Met Asp Ser Val Ala Phe Glu Asp Val Ser Val Ser Phe Ser Gln

1	5	10	15
Glu Glu Trp Ala Leu	Leu Ala Pro Ser	Gln Lys Lys Leu Tyr	Arg
20	25	30	
Asp Val Met Gln Glu	Thr Phe Lys Asn Leu	Ala Ser Ile Gly	Glu
35	40	45	
Lys Trp Glu Asp Pro	Asn Val Glu Asp	Gln His Lys Asn Gln	Gly
50	55	60	
Arg Asn Leu Arg Ser	His Thr Gly Glu	Arg Leu Cys Glu Gly	Lys
65	70	75	
Glu Gly Ser Gln Cys	Ala Glu Asn Phe Ser	Pro Asn Leu Ser	Val
80	85	90	
Thr Lys Lys Thr Ala	Gly Val Lys Pro Tyr	Glu Cys Thr Ile	Cys
95	100	105	
Gly Lys Ala Phe Met	Arg Leu Ser Ser Leu	Thr Arg His Met	Arg
110	115	120	
Ser His Thr Gly Tyr	Glu Leu Phe Glu Lys	Pro Tyr Lys Cys	Lys
125	130	135	
Glu Cys Glu Lys Ala	Phe Ser Tyr Leu Lys	Ser Phe Gln Arg	His
140	145	150	
Glu Arg Ser His Thr	Gly Glu Lys Pro Tyr	Lys Cys Lys Gln	Cys
155	160	165	
Gly Lys Thr Phe Ile	Tyr His Gln Pro Phe	Gln Arg His Glu	Arg
170	175	180	
Thr His Ile Gly Glu	Lys Pro Tyr Glu Cys	Lys Gln Cys Gly	Lys
185	190	195	
Ala Leu Ser Cys Ser	Ser Ser Ser Leu Arg	Val His Glu Arg	Ile
200	205	210	
Thr Gly Glu Lys Pro	Tyr Glu Cys Lys Gln	Cys Gly Lys Ala	Phe
215	220	225	
Ser Cys Ser Ser Ser	Ile Arg Val His Glu	Arg Thr His Thr	Gly
230	235	240	
Glu Lys Pro Tyr Ala	Cys Lys Glu Cys Gly	Lys Ala Phe Ile	Ser
245	250	255	
His Thr Ser Val Leu	Thr His Met Ile Thr	His Asn Gly Asp	Arg
260	265	270	
Pro Tyr Lys Cys Lys	Glu Cys Gly Lys Ala	Phe Ile Phe Pro	Ser
275	280	285	
Phe Leu Arg Val His	Glu Arg Ile His Thr	Gly Glu Lys Pro	Tyr
290	295	300	
Lys Cys Lys Gln Cys	Gly Lys Ala Phe Arg	Cys Ser Thr Ser	Ile
305	310	315	
Gln Ile His Glu Arg	Ile His Thr Gly Glu	Lys Pro Tyr Lys	Cys
320	325	330	
Lys Glu Cys Gly Lys	Ser Phe Ser Ala Arg	Pro Ala Phe Arg	Val
335	340	345	
His Val Arg Val His	Thr Gly Glu Lys Pro	Tyr Lys Cys Lys	Glu
350	355	360	
Cys Gly Lys Ala Phe	Ser Arg Ile Ser Tyr	Phe Arg Ile His	Glu
365	370	375	
Arg Thr His Thr Gly	Glu Lys Pro Tyr Glu	Cys Lys Lys Cys	Gly
380	385	390	
Lys Thr Phe Asn Tyr	Pro Leu Asp Leu Lys	Ile His Lys Arg	Asn
395	400	405	
His Thr Gly Glu Lys	Pro Tyr Glu Cys Lys	Glu Cys Ala Lys	Thr
410	415	420	
Phe Ile Ser Leu Glu	Asn Phe Arg Arg His	Met Ile Thr His	Thr

	425		430		435
Gly Asp Gly Pro Tyr Lys Cys Arg Asp Cys Gly Lys Val Phe Ile					
	440		445		450
Phe Pro Ser Ala Leu Arg Thr His Glu Arg Thr His Thr Gly Glu					
	455		460		465
Lys Pro Tyr Glu Cys Lys Gln Cys Gly Lys Ala Phe Ser Cys Ser					
	470		475		480
Ser Tyr Ile Arg Ile His Lys Arg Thr His Thr Gly Glu Lys Pro					
	485		490		495
Tyr Glu Cys Lys Glu Cys Gly Lys Ala Phe Ile Tyr Pro Thr Ser					
	500		505		510
Phe Gln Gly His Met Arg Met His Thr Gly Glu Lys Pro Tyr Lys					
	515		520		525
Cys Lys Glu Cys Gly Lys Ala Phe Ser Leu His Ser Ser Phe Gln					
	530		535		540
Arg His Thr Arg Ile His Asn Tyr Glu Lys Pro Leu Glu Cys Lys					
	545		550		555
Gln Cys Gly Lys Ala Phe Ser Val Ser Thr Ser Leu Lys Lys His					
	560		565		570
Met Arg Met His Asn Arg					
	575				

&lt;210&gt; 4

&lt;211&gt; 426

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3495982CD1

&lt;400&gt; 4

Met Arg Arg Asn Ser Ser Leu Ser Phe Gln Met Glu Arg Pro Leu			
1	5	10	15
Glu Glu Gln Val Gln Ser Lys Trp Ser Ser Ser Gln Gly Arg Thr			
	20	25	30
Gly Thr Gly Gly Ser Asp Val Leu Gln Met Gln Asn Ser Glu His			
	35	40	45
His Gly Gln Ser Ile Lys Thr Gln Thr Asp Ser Ile Ser Leu Glu			
	50	55	60
Asp Val Ala Val Asn Phe Thr Leu Glu Glu Trp Ala Leu Leu Asp			
	65	70	75
Pro Gly Gln Arg Asn Ile Tyr Arg Asp Val Met Arg Ala Thr Phe			
	80	85	90
Lys Asn Leu Ala Cys Ile Gly Glu Lys Trp Lys Asp Gln Asp Ile			
	95	100	105
Glu Asp Glu His Lys Asn Gln Gly Arg Asn Leu Arg Ser Pro Met			
	110	115	120
Val Glu Ala Leu Cys Glu Asn Lys Glu Asp Cys Pro Cys Gly Lys			
	125	130	135
Ser Thr Ser Gln Ile Pro Asp Leu Asn Thr Asn Leu Glu Thr Pro			
	140	145	150
Thr Gly Leu Lys Pro Cys Asp Cys Ser Val Cys Gly Glu Val Phe			
	155	160	165
Met His Gln Val Ser Leu Asn Arg His Met Arg Ser His Thr Glu			
	170	175	180

Gln Lys Pro Asn Glu Cys His Glu Tyr Gly Glu Lys Pro His Lys  
 185 190 195  
 Cys Lys Glu Cys Gly Lys Thr Phe Thr Arg Ser Ser Ser Ile Arg  
 200 205 210  
 Thr His Glu Arg Ile His Thr Gly Glu Lys Pro Tyr Glu Cys Lys  
 215 220 225  
 Glu Cys Gly Lys Ala Phe Ala Phe Leu Phe Ser Phe Arg Asn His  
 230 235 240  
 Ile Arg Ile His Thr Gly Glu Thr Pro Tyr Glu Cys Lys Glu Cys  
 245 250 255  
 Gly Lys Ala Phe Arg Tyr Leu Thr Ala Leu Arg Arg His Glu Lys  
 260 265 270  
 Asn His Thr Gly Glu Lys Pro Tyr Lys Cys Lys Gln Cys Gly Lys  
 275 280 285  
 Ala Phe Ile Tyr Thr Gln Pro Phe Leu Thr His Glu Arg Thr His  
 290 295 300  
 Thr Gly Glu Lys Pro Tyr Glu Cys Lys Gln Cys Gly Lys Ala Phe  
 305 310 315  
 Ser Cys Pro Thr Tyr Leu Arg Ser His Glu Lys Thr His Thr Gly  
 320 325 330  
 Glu Lys Pro Phe Val Cys Arg Glu Cys Gly Arg Ala Phe Phe Ser  
 335 340 345  
 His Ser Ser Leu Arg Lys His Val Ser His His Thr Arg Pro Pro  
 350 355 360  
 Val Leu Phe Phe Phe Phe Glu Thr Glu Ser Leu Pro Arg Leu Glu  
 365 370 375  
 Cys Ser Gly Ala Ile Ser Ala Tyr Cys Lys Leu Arg Leu Leu Gly  
 380 385 390  
 Ser Arg His Ser Pro Ala Ser Ala Ser Arg Val Ala Gly Thr Thr  
 395 400 405  
 Gly Ala Arg His His Ala Arg Leu Ile Phe Cys Ile Phe Ser Gly  
 410 415 420  
 Asp Gly Val Ser Pro Cys  
 425

&lt;210&gt; 5

&lt;211&gt; 786

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477891CD1

&lt;400&gt; 5

Met Ala Asn Asn Tyr Lys Lys Ile Val Leu Leu Lys Gly Leu Glu  
 1 5 10 15  
 Val Ile Asn Asp Tyr His Phe Arg Ile Val Lys Ser Leu Leu Ser  
 20 25 30  
 Asn Asp Leu Lys Leu Asn Pro Lys Met Lys Glu Glu Tyr Asp Lys  
 35 40 45  
 Ile Gln Ile Ala Asp Leu Met Glu Glu Lys Phe Pro Gly Asp Ala  
 50 55 60  
 Gly Leu Gly Lys Leu Ile Glu Phe Phe Lys Glu Ile Pro Thr Leu  
 65 70 75  
 Gly Asp Leu Ala Glu Thr Leu Lys Arg Glu Lys Leu Lys Val Lys

				80					85				90	
Gly	Ile	Ile	Pro	Ser	Lys	Lys	Thr	Lys	Gln	Lys	Glu	Val	Tyr	Pro
				95					100					105
Ala	Thr	Pro	Ala	Cys	Thr	Pro	Ser	Asn	Arg	Leu	Thr	Ala	Lys	Gly
				110					115					120
Ala	Glu	Glu	Thr	Leu	Gly	Pro	Gln	Lys	Arg	Lys	Lys	Pro	Ser	Glu
				125					130					135
Glu	Glu	Thr	Gly	Thr	Lys	Arg	Ser	Lys	Met	Ser	Lys	Glu	Gln	Thr
				140					145					150
Arg	Pro	Ser	Cys	Ser	Ala	Gly	Ala	Ser	Thr	Ser	Thr	Ala	Met	Gly
				155					160					165
Arg	Ser	Pro	Pro	Pro	Gln	Thr	Ser	Ser	Ser	Ala	Pro	Pro	Asn	Thr
				170					175					180
Ser	Ser	Thr	Glu	Ser	Leu	Lys	Pro	Leu	Ala	Asn	Arg	His	Ala	Thr
				185					190					195
Ala	Ser	Lys	Asn	Ile	Phe	Arg	Glu	Asp	Pro	Ile	Ile	Ala	Met	Val
				200					205					210
Leu	Asn	Ala	Thr	Lys	Val	Phe	Lys	Tyr	Glu	Ser	Ser	Glu	Asn	Glu
				215					220					225
Gln	Arg	Arg	Met	Phe	His	Ala	Thr	Val	Ala	Thr	Gln	Thr	Gln	Phe
				230					235					240
Phe	His	Val	Lys	Val	Leu	Asn	Ile	Asn	Leu	Lys	Arg	Lys	Phe	Ile
				245					250					255
Lys	Lys	Arg	Ile	Ile	Ile	Ile	Ser	Asn	Tyr	Ser	Lys	Arg	Asn	Ser
				260					265					270
Leu	Leu	Glu	Val	Asn	Glu	Ala	Ser	Ser	Val	Ser	Glu	Ala	Gly	Pro
				275					280					285
Asp	Gln	Thr	Phe	Glu	Val	Pro	Lys	Asp	Ile	Ile	Arg	Arg	Ala	Lys
				290					295					300
Lys	Ile	Pro	Lys	Ile	Asn	Ile	Leu	His	Lys	Gln	Thr	Ser	Gly	Tyr
				305					310					315
Ile	Val	Tyr	Gly	Leu	Phe	Met	Leu	His	Thr	Lys	Ile	Val	Asn	Arg
				320					325					330
Lys	Thr	Thr	Ile	Tyr	Glu	Ile	Gln	Asp	Lys	Thr	Gly	Ser	Met	Ala
				335					340					345
Val	Val	Gly	Lys	Gly	Glu	Cys	His	Asn	Ile	Pro	Cys	Glu	Lys	Gly
				350					355					360
Asp	Lys	Leu	Arg	Leu	Phe	Cys	Phe	Arg	Leu	Arg	Lys	Arg	Glu	Asn
				365					370					375
Met	Ser	Lys	Leu	Met	Ser	Glu	Met	His	Ser	Phe	Ile	Gln	Ile	Gln
				380					385					390
Lys	Asn	Thr	Asn	Gln	Arg	Ser	His	Asp	Ser	Arg	Ser	Met	Ala	Leu
				395					400					405
Pro	Gln	Glu	Gln	Ser	Gln	His	Pro	Lys	Pro	Ser	Glu	Ala	Ser	Thr
				410					415					420
Thr	Leu	Pro	Glu	Ser	His	Leu	Lys	Thr	Pro	Gln	Met	Pro	Pro	Thr
				425					430					435
Thr	Pro	Ser	Ser	Ser	Phe	Phe	Thr	Lys	Lys	Ser	Glu	Asp	Thr	Ile
				440					445					450
Ser	Lys	Met	Asn	Asp	Phe	Met	Arg	Met	Gln	Ile	Leu	Lys	Glu	Gly
				455					460					465
Ser	His	Phe	Pro	Gly	Pro	Phe	Met	Thr	Ser	Ile	Gly	Pro	Ala	Glu
				470					475					480
Ser	His	Pro	His	Thr	Pro	Gln	Met	Pro	Pro	Ser	Thr	Pro	Ser	Ser
				485					490					495
Ser	Phe	Leu	Thr	Thr	Lys	Ser	Glu	Asp	Thr	Ile	Ser	Lys	Met	Asn

	500		505		510
Asp Phe Met Arg	Met Gln Ile Leu Lys	Glu Gly Ser His Phe	Pro		
	515		520		525
Gly Pro Phe Met	Thr Ser Ile Gly Pro	Ala Glu Ser His Pro	His		
	530		535		540
Thr Pro Gln Met	Pro Pro Ser Thr Pro	Ser Ser Ser Phe Leu	Thr		
	545		550		555
Thr Leu Lys Pro	Arg Leu Lys Thr Glu	Pro Glu Glu Val Ser	Ile		
	560		565		570
Glu Asp Ser Ala	Gln Ser Asp Leu Lys	Glu Val Met Val Leu	Asn		
	575		580		585
Ala Thr Glu Ser	Phe Val Tyr Glu Pro	Lys Glu Gln Lys Lys	Met		
	590		595		600
Phe His Ala Thr	Val Ala Thr Glu Asn	Glu Val Phe Arg Val	Lys		
	605		610		615
Val Phe Asn Ile	Asp Leu Lys Glu Lys	Phe Thr Pro Lys Lys	Ile		
	620		625		630
Ile Ala Ile Ala	Asn Tyr Val Cys Arg	Asn Gly Phe Leu Glu	Val		
	635		640		645
Tyr Pro Phe Thr	Leu Val Ala Asp Val	Asn Ala Asp Arg Asn	Met		
	650		655		660
Glu Ile Pro Lys	Gly Leu Ile Arg Ser	Ala Ser Val Thr Pro	Lys		
	665		670		675
Ile Asn Gln Leu	Cys Ser Gln Thr Lys	Gly Ser Phe Val Asn	Gly		
	680		685		690
Val Phe Glu Val	His Lys Lys Asn Val	Arg Gly Glu Phe Thr	Tyr		
	695		700		705
Tyr Glu Ile Gln	Asp Asn Thr Gly Lys	Met Glu Val Val Val	His		
	710		715		720
Gly Arg Leu Thr	Thr Ile Asn Cys Glu	Glu Gly Asp Lys Leu	Lys		
	725		730		735
Leu Thr Cys Phe	Glu Leu Ala Pro Lys	Ser Gly Asn Thr Gly	Glu		
	740		745		750
Leu Arg Ser Val	Ile His Ser His Ile	Lys Val Ile Lys Thr	Arg		
	755		760		765
Lys Asn Lys Lys	Asp Ile Leu Asn Pro	Asp Ser Ser Met Glu	Thr		
	770		775		780
Ser Pro Asp Phe	Phe Phe				
	785				

&lt;210&gt; 6

&lt;211&gt; 617

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 72688352CD1

&lt;400&gt; 6

Met Ile Lys Ser Gln	Glu Ser Leu Thr	Leu Glu Asp Val	Ala Val
1	5	10	15
Glu Phe Thr Trp	Glu Glu Trp Gln	Leu Leu Gly Pro	Ala Gln Lys
	20	25	30
Asp Leu Tyr Arg	Asp Val Met Leu	Glu Asn Tyr Ser	Asn Leu Val
	35	40	45

Ser Val Gly Tyr Gln Ala Ser Lys Pro Asp Ala Leu Phe Lys Leu		
	50	55 60
Glu Gln Gly Glu Pro Trp Thr Val Glu Asn Glu Ile His Ser Gln		
	65	70 75
Ile Cys Pro Glu Ile Lys Lys Val Asp Asn His Leu Gln Met His		
	80	85 90
Ser Gln Lys Gln Arg Cys Leu Lys Arg Val Glu Gln Cys His Lys		
	95	100 105
His Asn Ala Phe Gly Asn Ile Ile His Gln Arg Lys Ser Asp Phe		
	110	115 120
Pro Leu Arg Gln Asn His Asp Thr Phe Asp Leu His Gly Lys Ile		
	125	130 135
Leu Lys Ser Asn Leu Ser Leu Val Asn Gln Asn Lys Arg Tyr Glu		
	140	145 150
Ile Lys Asn Ser Val Gly Val Asn Gly Asp Gly Lys Ser Phe Leu		
	155	160 165
His Ala Lys His Glu Gln Phe His Asn Glu Met Asn Phe Pro Glu		
	170	175 180
Gly Gly Asn Ser Val Asn Thr Asn Ser Gln Phe Ile Lys His Gln		
	185	190 195
Arg Thr Gln Asn Ile Asp Lys Pro His Val Cys Thr Glu Cys Gly		
	200	205 210
Lys Ala Phe Leu Lys Lys Ser Arg Leu Ile Tyr His Gln Arg Val		
	215	220 225
His Thr Gly Glu Lys Pro His Gly Cys Ser Ile Cys Gly Lys Ala		
	230	235 240
Phe Ser Arg Lys Ser Gly Leu Thr Glu His Gln Arg Asn His Thr		
	245	250 255
Gly Glu Lys Pro Tyr Glu Cys Thr Glu Cys Asp Lys Ala Phe Arg		
	260	265 270
Trp Lys Ser Gln Leu Asn Ala His Gln Lys Ile His Thr Gly Glu		
	275	280 285
Lys Ser Tyr Ile Cys Ser Asp Cys Gly Lys Gly Phe Ile Lys Lys		
	290	295 300
Ser Arg Leu Ile Asn His Gln Arg Val His Thr Gly Glu Lys Pro		
	305	310 315
His Gly Cys Ser Leu Cys Gly Lys Ala Phe Ser Lys Arg Ser Arg		
	320	325 330
Leu Thr Glu His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Glu		
	335	340 345
Cys Thr Glu Cys Asp Lys Ala Phe Arg Trp Lys Ser Gln Leu Asn		
	350	355 360
Ala His Gln Lys Ala His Thr Gly Glu Lys Ser Tyr Ile Cys Arg		
	365	370 375
Asp Cys Gly Lys Gly Phe Ile Gln Lys Gly Asn Leu Ile Val His		
	380	385 390
Gln Arg Ile His Thr Gly Glu Lys Pro Tyr Ile Cys Asn Glu Cys		
	395	400 405
Gly Lys Gly Phe Ile Gln Lys Gly Asn Leu Leu Ile His Arg Arg		
	410	415 420
Thr His Thr Gly Glu Lys Pro Tyr Val Cys Asn Glu Cys Gly Lys		
	425	430 435
Gly Phe Ser Gln Lys Thr Cys Leu Ile Ser His Gln Arg Phe His		
	440	445 450
Thr Gly Lys Thr Pro Phe Val Cys Thr Glu Cys Gly Lys Ser Cys		
	455	460 465

```

Ser His Lys Ser Gly Leu Ile Asn His Gln Arg Ile His Thr Gly
      470                      475                      480
Glu Lys Pro Tyr Thr Cys Ser Asp Cys Gly Lys Ala Phe Arg Asp
      485                      490                      495
Lys Ser Cys Leu Asn Arg His Arg Arg Thr His Thr Gly Glu Arg
      500                      505                      510
Pro Tyr Gly Cys Ser Asp Cys Gly Lys Ala Phe Ser His Leu Ser
      515                      520                      525
Cys Leu Val Tyr His Lys Gly Met Leu His Ala Arg Glu Lys Cys
      530                      535                      540
Val Gly Ser Val Lys Leu Glu Asn Pro Cys Ser Glu Ser His Ser
      545                      550                      555
Leu Ser His Thr Arg Asp Leu Ile Gln Asp Lys Asp Ser Val Asn
      560                      565                      570
Met Val Thr Leu Gln Met Pro Ser Val Ala Ala Gln Thr Ser Leu
      575                      580                      585
Thr Asn Ser Ala Phe Gln Ala Glu Ser Lys Val Ala Ile Val Ser
      590                      595                      600
Gln Pro Val Ala Arg Ser Ser Val Ser Ala Asp Ser Arg Ile Cys
      605                      610                      615
Thr Glu

```

&lt;210&gt; 7

&lt;211&gt; 249

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7490652CD1

&lt;400&gt; 7

```

Met Ala Val Gly Lys Asn Lys His Leu Met Lys Gly Gly Lys Lys
  1                      5                      10                      15
Gly Ala Glu Asn Arg Val Val Asp Pro Phe Ser Lys Lys Asp Trp
      20                      25                      30
Cys Asp Val Lys Ala Leu Ala Met Phe Asn Ile Arg Asn Ile Gly
      35                      40                      45
Glu Thr Leu Val Thr Arg Thr Arg Gly Thr Lys Ile Ala Ser Asp
      50                      55                      60
Ser Leu Lys Arg Arg Val Phe Glu Val Ser Leu Ala Asp Leu Gln
      65                      70                      75
Asn Asp Glu Val Ala Phe Arg Lys Phe Lys Leu Ile Ala Glu Asp
      80                      85                      90
Val Gln Lys Lys Thr Asn Phe Gln Gly Met Asp Leu Pro Asp Glu
      95                      100                     105
Met Cys Ser Val Val Lys Lys Trp Gln Thr Met Ile Glu Pro His
      110                     115                     120
Ile Asp Val Lys Thr Thr Asp Gly Tyr Leu Phe His Leu Leu Cys
      125                     130                     135
Asp Phe Thr Lys Lys His Asn Leu Ile Gln Lys Ala Ser Tyr Ala
      140                     145                     150
Gln His Gln Gln Val Cys Glu Ile Gln Lys Lys Met Met Glu Ile
      155                     160                     165
Met Thr Lys Gly Ala Asn Asp Leu Lys Glu Val Val Asn Lys Leu

```

	170		175		180
Ile Pro Gly Ser Thr Gly Lys Glu Lys		Leu Cys Leu Ser Ile Tyr			
	185		190		195
Leu Leu His Asp Val Phe Val Arg Lys		Val Lys Met Leu Lys Met			
	200		205		210
Pro Lys Phe Asp Leu Gly Lys Phe Met		Gly Asn Cys Ser Gly Lys			
	215		220		225
Ala Thr Gly Asp Glu Thr Gly Ala Lys		Val Glu Leu Ala Asp Gly			
	230		235		240
Tyr Glu Ala Leu Val Gln Glu Ser Val					
	245				

&lt;210&gt; 8

&lt;211&gt; 384

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7489744CD1

&lt;400&gt; 8

Met Glu Val Ile Val Glu Asn Leu His Leu Pro Thr Ser Pro Ile			
1	5	10	15
Pro Pro Val Ala Gly Ala Glu Ser Gly Pro Gln Arg Ala Leu Ser			
	20	25	30
Ser Pro Thr Ala Ala Ala Gly Leu Val Thr Ile Thr Pro Arg Glu			
	35	40	45
Glu Pro Gln Leu Pro Gln Pro Ala Pro Val Thr Ile Thr Ala Thr			
	50	55	60
Met Ser Ser Glu Ala Glu Thr Gln Gln Pro Pro Ala Ala Pro Pro			
	65	70	75
Ala Ala Pro Ala Leu Ser Ala Ala Asp Thr Lys Pro Gly Thr Thr			
	80	85	90
Gly Ser Gly Ala Gly Ser Gly Gly Pro Gly Gly Leu Thr Ser Ala			
	95	100	105
Ala Pro Ala Gly Gly Asp Lys Lys Val Ile Ala Thr Lys Val Leu			
	110	115	120
Gly Thr Val Lys Trp Phe Asn Val Arg Asn Gly Tyr Gly Phe Ile			
	125	130	135
Asn Arg Asn Asp Thr Lys Glu Asp Val Phe Val His Gln Thr Ala			
	140	145	150
Ile Lys Lys Asn Asn Pro Arg Lys Tyr Leu Arg Ser Val Gly Asp			
	155	160	165
Gly Glu Thr Val Glu Phe Asp Val Val Glu Gly Glu Lys Gly Ala			
	170	175	180
Glu Ala Ala Asn Val Thr Gly Pro Gly Gly Val Pro Val Gln Gly			
	185	190	195
Ser Lys Tyr Ala Ala Asp Arg Asn His Tyr Arg Arg Tyr Pro Arg			
	200	205	210
Arg Arg Gly Pro Pro Arg Asn Tyr Gln Gln Asn Tyr Gln Asn Ser			
	215	220	225
Glu Ser Gly Glu Lys Asn Glu Gly Ser Glu Ser Ala Pro Glu Gly			
	230	235	240
Gln Ala Gln Gln Arg Arg Pro Tyr Arg Arg Arg Arg Phe Pro Pro			
	245	250	255

```

Tyr Tyr Met Arg Arg Pro Tyr Gly Arg Arg Pro Gln Tyr Ser Asn
      260                      265                      270
Pro Pro Val Gln Gly Glu Val Met Glu Gly Ala Asp Asn Gln Gly
      275                      280                      285
Ala Gly Glu Gln Gly Arg Pro Val Arg Gln Asn Met Tyr Arg Gly
      290                      295                      300
Tyr Arg Pro Arg Phe Arg Arg Gly Pro Pro Arg Gln Arg Gln Pro
      305                      310                      315
Arg Glu Asp Gly Asn Glu Glu Asp Lys Glu Asn Gln Gly Asp Glu
      320                      325                      330
Thr Gln Gly Gln Gln Pro Pro Gln Arg Arg Tyr Arg Arg Asn Phe
      335                      340                      345
Asn Tyr Arg Arg Arg Arg Pro Glu Asn Pro Lys Pro Gln Asp Gly
      350                      355                      360
Gln Glu Thr Lys Ala Ala Asp Pro Pro Ala Glu Asn Ser Ser Ala
      365                      370                      375
Pro Glu Ala Glu Gln Gly Gly Ala Glu
      380

```

&lt;210&gt; 9

&lt;211&gt; 312

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3363382CD1

&lt;400&gt; 9

```

Met Ala Asp Gly Asp Ser Gly Ser Glu Arg Gly Gly Gly Gly Gly
  1                      5                      10                      15
Pro Cys Gly Phe Gln Pro Ala Ser Arg Gly Gly Gly Glu Gln Glu
      20                      25                      30
Thr Gln Glu Leu Ala Ser Lys Arg Leu Asp Ile Gln Asn Lys Arg
      35                      40                      45
Phe Tyr Leu Asp Val Lys Gln Asn Ala Lys Gly Arg Phe Leu Lys
      50                      55                      60
Ile Ala Glu Val Gly Ala Gly Gly Ser Lys Ser Arg Leu Thr Leu
      65                      70                      75
Ser Met Ala Val Ala Ala Glu Phe Arg Asp Ser Leu Gly Asp Phe
      80                      85                      90
Ile Glu His Tyr Ala Gln Leu Gly Pro Ser Ser Pro Glu Gln Leu
      95                      100                     105
Ala Ala Gly Ala Glu Glu Gly Gly Gly Pro Arg Arg Ala Leu Lys
      110                     115                     120
Ser Glu Phe Leu Val Arg Glu Asn Arg Lys Tyr Tyr Leu Asp Leu
      125                     130                     135
Lys Glu Asn Gln Arg Gly Arg Phe Leu Arg Ile Arg Gln Thr Val
      140                     145                     150
Asn Arg Gly Gly Gly Gly Phe Gly Ala Gly Pro Gly Pro Gly Gly
      155                     160                     165
Leu Gln Ser Gly Gln Thr Ile Ala Leu Pro Ala Gln Gly Leu Ile
      170                     175                     180
Glu Phe Arg Asp Ala Leu Ala Lys Leu Ile Asp Asp Tyr Gly Gly
      185                     190                     195
Glu Asp Asp Glu Leu Ala Gly Gly Pro Gly Gly Gly Ala Gly Gly

```

	200		205		210
Pro Gly Gly Gly	Leu Tyr Gly Glu Leu	Pro Glu Gly Thr Ser	Ile		
	215		220		225
Thr Val Asp Ser	Lys Arg Phe Phe Phe	Asp Val Gly Cys Asn	Lys		
	230		235		240
Tyr Gly Val Phe	Leu Arg Val Ser Glu	Val Lys Pro Ser Tyr	Arg		
	245		250		255
Asn Ala Ile Thr	Val Pro Phe Lys Ala	Trp Gly Lys Phe Gly	Gly		
	260		265		270
Ala Phe Cys Arg	Tyr Ala Asp Glu Met	Lys Glu Ile Gln Glu	Arg		
	275		280		285
Gln Arg Asp Lys	Leu Tyr Glu Arg Arg	Gly Gly Gly Ser Gly	Gly		
	290		295		300
Gly Glu Glu Ser	Glu Gly Glu Glu Val	Asp Glu Asp			
	305		310		

&lt;210&gt; 10

&lt;211&gt; 441

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7491148CD1

&lt;400&gt; 10

Met Lys Asp His Asp	Ala Ile Lys Leu	Phe Val Gly Gln Ile	Pro
1	5	10	15
Arg Gly Leu Asp Glu	Gln Asp Leu Lys	Pro Leu Phe Glu Glu	Phe
	20	25	30
Gly Arg Ile Tyr Glu	Leu Thr Val Leu	Lys Asp Arg Leu Thr	Gly
	35	40	45
Leu His Lys Gly Cys	Ala Phe Leu Thr	Tyr Cys Ala Arg Asp	Ser
	50	55	60
Ala Leu Lys Ala Gln	Ser Ala Leu His	Glu Gln Lys Thr Leu	Pro
	65	70	75
Gly Phe His Ile Leu	Asn Asn Asn Asn	Asn Asn Lys Asn Arg	Pro
	80	85	90
Glu Asp Arg Lys Leu	Phe Val Gly Met	Leu Gly Lys Gln Gln	Gly
	95	100	105
Glu Glu Asp Val Arg	Arg Leu Phe Gln	Pro Phe Gly His Ile	Glu
	110	115	120
Glu Cys Thr Val Leu	Arg Ser Pro Asp	Gly Thr Ser Lys Gly	Cys
	125	130	135
Ala Phe Val Lys Phe	Gly Ser Gln Gly	Glu Ala Gln Ala Ala	Ile
	140	145	150
Arg Gly Leu His Gly	Ser Arg Thr Met	Ala Gly Ala Ser Ser	Ser
	155	160	165
Leu Val Val Lys Leu	Ala Asp Thr Asp	Arg Glu Arg Ala Leu	Arg
	170	175	180
Arg Met Gln Gln Met	Ala Gly His Leu	Gly Ala Phe His Pro	Ala
	185	190	195
Pro Leu Pro Leu Gly	Ala Cys Gly Ala	Tyr Thr Thr Ala Ile	Leu
	200	205	210
Gln His Gln Ala Ala	Leu Leu Ala Ala	Ala Gln Gly Pro Gly	Leu
	215	220	225

Gly Pro Val Ala Ala Val Ala Ala Gln Met Gln His Val Ala Ala  
 230 235 240  
 Phe Ser Leu Val Ala Ala Pro Leu Leu Pro Ala Ala Ala Ala Asn  
 245 250 255  
 Ser Pro Pro Gly Ser Gly Pro Gly Thr Leu Pro Gly Leu Pro Ala  
 260 265 270  
 Pro Ile Gly Val Asn Gly Val Arg Pro Ser Asp Thr Pro Arg Ser  
 275 280 285  
 Asn Gly Gln Pro Gly Ser Asp Thr Leu Tyr Asn Asn Gly Leu Ser  
 290 295 300  
 Pro Tyr Pro Ala Gln Ser Pro Gly Val Ala Asp Pro Leu Gln Gln  
 305 310 315  
 Ala Tyr Ala Gly Met His His Tyr Ala Ala Ala Tyr Pro Ser Ala  
 320 325 330  
 Tyr Ala Pro Val Ser Thr Ala Phe Pro Gln Gln Pro Ser Ala Leu  
 335 340 345  
 Pro Gln Gln Gln Arg Glu Gly Pro Glu Gly Cys Asn Leu Phe Ile  
 350 355 360  
 Tyr His Leu Pro Gln Glu Phe Gly Asp Ala Glu Leu Ile Gln Thr  
 365 370 375  
 Phe Leu Pro Phe Gly Ala Val Val Ser Ala Lys Val Phe Val Asp  
 380 385 390  
 Arg Ala Thr Asn Gln Ser Lys Cys Phe Gly Phe Val Ser Phe Asp  
 395 400 405  
 Asn Pro Thr Ser Ala Gln Thr Ala Ile Gln Ala Met Asn Gly Phe  
 410 415 420  
 Gln Ile Gly Met Lys Arg Leu Lys Val Gln Leu Lys Arg Pro Lys  
 425 430 435  
 Asp Ala Asn Arg Pro Tyr  
 440

&lt;210&gt; 11

&lt;211&gt; 493

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 8126343CD1

&lt;400&gt; 11

Met Ala Thr Asp Leu Pro Ile Met Ala Arg Gly Pro Ala Arg Ser  
 1 5 10 15  
 Ala Ala Pro Ala Gly Gly Ser Ser Ser Gly Cys Gly Ala Arg Gln  
 20 25 30  
 Gly Arg Ala Gly Gly Val Leu Ala Met Ala Gly Leu Ser Asp  
 35 40 45  
 Leu Glu Leu Arg Arg Glu Leu Gln Ala Leu Gly Phe Gln Pro Gly  
 50 55 60  
 Pro Ile Thr Asp Thr Thr Arg Asp Val Tyr Arg Asn Lys Leu Arg  
 65 70 75  
 Arg Leu Arg Gly Glu Ala Arg Leu Arg Asp Glu Glu Arg Leu Arg  
 80 85 90  
 Glu Glu Ala Arg Pro Arg Gly Glu Glu Arg Leu Arg Glu Glu Ala  
 95 100 105  
 Arg Leu Arg Glu Asp Ala Pro Leu Arg Ala Arg Pro Ala Ala Ala

	110		115		120
Ser Pro Arg Ala	Glu Pro Trp Leu Ser	Gln Pro Ala Ser Gly Ser			
	125		130		135
Ala Tyr Ala Thr	Pro Gly Ala Tyr Gly	Asp Ile Arg Pro Ser Ala			
	140		145		150
Ala Ser Trp Val	Gly Ser Arg Gly Leu	Ala Tyr Pro Ala Arg Pro			
	155		160		165
Ala Gln Leu Arg	Arg Arg Ala Ser Val	Arg Gly Ser Ser Glu Glu			
	170		175		180
Asp Glu Asp Ala	Arg Thr Pro Asp Arg	Ala Thr Gln Gly Pro Gly			
	185		190		195
Leu Ala Ala Arg	Arg Trp Trp Ala Ala	Ser Pro Ala Pro Ala Arg			
	200		205		210
Leu Pro Ser Ser	Leu Leu Gly Pro Asp	Pro Arg Pro Gly Leu Arg			
	215		220		225
Ala Thr Arg Ala	Gly Pro Ala Gly Ala	Ala Arg Ala Arg Pro Glu			
	230		235		240
Val Gly Arg Arg	Leu Glu Arg Trp Leu	Ser Arg Leu Leu Leu Trp			
	245		250		255
Ala Ser Leu Gly	Leu Leu Leu Val Phe	Leu Gly Ile Leu Trp Val			
	260		265		270
Lys Met Gly Lys	Pro Ser Ala Pro Gln	Glu Ala Glu Asp Asn Met			
	275		280		285
Lys Leu Leu Pro	Val Asp Cys Glu Arg	Lys Thr Asp Glu Phe Cys			
	290		295		300
Gln Ala Lys Gln	Lys Ala Ala Leu Leu	Glu Leu Leu His Glu Leu			
	305		310		315
Tyr Asn Phe Leu	Ala Ile Gln Ala Gly	Asn Phe Glu Cys Gly Asn			
	320		325		330
Pro Glu Asn Leu	Lys Ser Lys Cys Ile	Pro Val Met Glu Ala Gln			
	335		340		345
Glu Tyr Ile Ala	Asn Val Thr Ser Ser	Ser Ser Ala Lys Phe Glu			
	350		355		360
Ala Ala Leu Thr	Trp Ile Leu Ser Ser	Asn Lys Asp Val Gly Ile			
	365		370		375
Trp Leu Lys Gly	Glu Asp Gln Ser Glu	Leu Val Thr Thr Val Asp			
	380		385		390
Lys Val Val Cys	Leu Glu Ser Ala His	Pro Arg Met Gly Val Gly			
	395		400		405
Cys Arg Leu Ser	Arg Ala Leu Leu Thr	Ala Val Thr Asn Val Leu			
	410		415		420
Ile Phe Phe Trp	Cys Leu Ala Phe Leu	Trp Gly Leu Leu Ile Leu			
	425		430		435
Leu Lys Tyr Arg	Trp Arg Lys Leu Glu	Glu Glu Glu Gln Ala Met			
	440		445		450
Tyr Glu Met Val	Lys Lys Ile Ile Asp	Val Val Gln Asp His Tyr			
	455		460		465
Val Asp Trp Glu	Gln Asp Met Glu Arg	Tyr Pro Tyr Val Gly Ile			
	470		475		480
Leu His Val Arg	Asp Ser Leu Ile Pro	Pro Gln Ser Arg			
	485		490		

&lt;210&gt; 12

&lt;211&gt; 553

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7044055CD1

&lt;400&gt; 12

Met	Ala	Ala	Val	Ser	Leu	Arg	Leu	Gly	Asp	Leu	Val	Trp	Gly	Lys
1				5					10					15
Leu	Gly	Arg	Tyr	Pro	Pro	Trp	Pro	Gly	Lys	Ile	Val	Asn	Pro	Pro
				20					25					30
Lys	Asp	Leu	Lys	Lys	Pro	Arg	Gly	Lys	Lys	Cys	Phe	Phe	Val	Lys
				35					40					45
Phe	Phe	Gly	Thr	Glu	Asp	His	Ala	Trp	Ile	Lys	Val	Glu	Gln	Leu
				50					55					60
Lys	Pro	Tyr	His	Ala	His	Lys	Glu	Glu	Met	Ile	Lys	Ile	Asn	Lys
				65					70					75
Gly	Lys	Arg	Phe	Gln	Gln	Ala	Val	Asp	Ala	Val	Glu	Glu	Phe	Leu
				80					85					90
Arg	Arg	Ala	Lys	Gly	Lys	Asp	Gln	Thr	Ser	Ser	His	Asn	Ser	Ser
				95					100					105
Asp	Asp	Lys	Asn	Arg	Arg	Asn	Ser	Ser	Glu	Glu	Arg	Ser	Arg	Pro
				110					115					120
Asn	Ser	Gly	Asp	Glu	Lys	Arg	Lys	Leu	Ser	Leu	Ser	Glu	Gly	Lys
				125					130					135
Val	Lys	Lys	Asn	Met	Gly	Glu	Gly	Lys	Lys	Arg	Val	Ser	Ser	Gly
				140					145					150
Ser	Ser	Glu	Arg	Gly	Ser	Lys	Ser	Pro	Leu	Lys	Arg	Ala	Gln	Glu
				155					160					165
Gln	Ser	Pro	Arg	Lys	Arg	Gly	Arg	Pro	Pro	Lys	Asp	Glu	Lys	Asp
				170					175					180
Leu	Thr	Ile	Pro	Glu	Ser	Ser	Thr	Val	Lys	Gly	Met	Met	Ala	Gly
				185					190					195
Pro	Met	Ala	Ala	Phe	Lys	Trp	Gln	Pro	Thr	Ala	Ser	Glu	Pro	Val
				200					205					210
Lys	Asp	Ala	Asp	Pro	His	Phe	His	His	Phe	Leu	Leu	Ser	Gln	Thr
				215					220					225
Glu	Lys	Pro	Ala	Val	Cys	Tyr	Gln	Ala	Ile	Thr	Lys	Lys	Leu	Lys
				230					235					240
Ile	Cys	Glu	Glu	Glu	Thr	Gly	Ser	Thr	Ser	Ile	Gln	Ala	Ala	Asp
				245					250					255
Ser	Thr	Ala	Val	Asn	Gly	Ser	Ile	Thr	Pro	Thr	Asp	Lys	Lys	Ile
				260					265					270
Gly	Phe	Leu	Gly	Leu	Gly	Leu	Met	Gly	Ser	Gly	Ile	Val	Ser	Asn
				275					280					285
Leu	Leu	Lys	Met	Gly	His	Thr	Val	Thr	Val	Trp	Asn	Arg	Thr	Ala
				290					295					300
Glu	Lys	Cys	Asp	Leu	Phe	Ile	Gln	Glu	Gly	Ala	Arg	Leu	Gly	Arg
				305					310					315
Thr	Pro	Ala	Glu	Val	Val	Ser	Thr	Cys	Asp	Ile	Thr	Phe	Ala	Cys
				320					325					330
Val	Ser	Asp	Pro	Lys	Ala	Ala	Lys	Asp	Leu	Val	Leu	Gly	Pro	Ser
				335					340					345
Gly	Val	Leu	Gln	Gly	Ile	Arg	Pro	Gly	Lys	Cys	Tyr	Val	Asp	Met
				350					355					360
Ser	Thr	Val	Asp	Ala	Asp	Thr	Val	Thr	Glu	Leu	Ala	Gln	Val	Ile
				365					370					375
Val	Ser	Arg	Gly	Gly	Arg	Phe	Leu	Glu	Ala	Pro	Val	Ser	Gly	Asn

	380		385		390
Gln Gln Leu Ser Asn Asp Gly Met Leu Val Ile Leu Ala Ala Gly					
	395		400		405
Asp Arg Gly Leu Tyr Glu Asp Cys Ser Ser Cys Phe Gln Ala Met					
	410		415		420
Gly Lys Thr Ser Phe Phe Leu Gly Glu Val Gly Asn Ala Ala Lys					
	425		430		435
Met Met Leu Ile Val Asn Met Val Gln Gly Ser Phe Met Ala Thr					
	440		445		450
Ile Ala Glu Gly Leu Thr Leu Ala Gln Val Thr Gly Gln Ser Gln					
	455		460		465
Gln Thr Leu Leu Asp Ile Leu Asn Gln Gly Gln Leu Ala Ser Ile					
	470		475		480
Phe Leu Asp Gln Lys Cys Gln Asn Ile Leu Gln Gly Asn Phe Lys					
	485		490		495
Pro Asp Phe Tyr Leu Lys Tyr Ile Gln Lys Asp Leu Arg Leu Ala					
	500		505		510
Ile Ala Leu Gly Asp Ala Val Asn His Pro Thr Pro Met Ala Ala					
	515		520		525
Ala Ala Asn Glu Val Tyr Lys Arg Ala Lys Ala Leu Asp Gln Ser					
	530		535		540
Asp Asn Asp Met Ser Ala Val Tyr Arg Ala Tyr Ile His					
	545		550		

&lt;210&gt; 13

&lt;211&gt; 1726

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7493424CD1

&lt;400&gt; 13

Met Lys Ala Gln Lys Ser Gly Lys Glu Gln Gln Leu Asp Ile Met		
1	5	10 15
Asn Lys Gln Tyr Gln Gln Leu Glu Ser Arg Leu Asp Glu Ile Leu		
	20	25 30
Ser Arg Ile Ala Lys Glu Thr Glu Glu Ile Lys Asp Leu Glu Glu		
	35	40 45
Gln Leu Thr Glu Gly Gln Ile Ala Ala Asn Glu Ala Leu Lys Lys		
	50	55 60
Asp Leu Glu Gly Val Ile Ser Gly Leu Gln Glu Tyr Leu Gly Thr		
	65	70 75
Ile Lys Gly Gln Ala Thr Gln Ala Gln Asn Glu Cys Arg Lys Leu		
	80	85 90
Arg Asp Glu Lys Glu Thr Leu Leu Gln Arg Leu Thr Glu Val Glu		
	95	100 105
Gln Glu Arg Asp Gln Leu Glu Ile Val Ala Met Asp Ala Glu Asn		
	110	115 120
Met Arg Lys Glu Leu Ala Glu Leu Glu Ser Ala Leu Gln Glu Gln		
	125	130 135
His Glu Val Asn Ala Ser Leu Gln Gln Thr Gln Gly Asp Leu Ser		
	140	145 150
Ala Tyr Glu Ala Glu Leu Glu Ala Arg Leu Asn Leu Arg Asp Ala		
	155	160 165

Glu	Ala	Asn	Gln	Leu	Lys	Glu	Glu	Leu	Glu	Lys	Val	Thr	Arg	Leu	170	175	180
Thr	Gln	Leu	Glu	Gln	Ser	Ala	Leu	Gln	Ala	Glu	Leu	Glu	Lys	Glu	185	190	195
Arg	Gln	Ala	Leu	Lys	Asn	Ala	Leu	Gly	Lys	Ala	Gln	Phe	Ser	Glu	200	205	210
Glu	Lys	Glu	Gln	Glu	Asn	Ser	Glu	Leu	His	Ala	Lys	Leu	Lys	His	215	220	225
Leu	Gln	Asp	Asp	Asn	Asn	Leu	Leu	Lys	Gln	Gln	Leu	Lys	Asp	Phe	230	235	240
Gln	Asn	His	Leu	Asn	His	Val	Val	Asp	Gly	Leu	Val	Arg	Pro	Glu	245	250	255
Glu	Val	Ala	Ala	Arg	Val	Asp	Glu	Leu	Arg	Arg	Lys	Leu	Lys	Leu	260	265	270
Gly	Thr	Gly	Glu	Met	Asn	Ile	His	Ser	Pro	Ser	Asp	Val	Leu	Gly	275	280	285
Lys	Ser	Leu	Ala	Asp	Leu	Gln	Lys	Gln	Phe	Ser	Glu	Ile	Leu	Ala	290	295	300
Arg	Ser	Lys	Trp	Glu	Arg	Asp	Glu	Ala	Gln	Val	Arg	Glu	Arg	Lys	305	310	315
Leu	Gln	Glu	Glu	Met	Ala	Leu	Gln	Gln	Glu	Lys	Leu	Ala	Thr	Gly	320	325	330
Gln	Glu	Glu	Phe	Arg	Gln	Ala	Cys	Glu	Arg	Ala	Leu	Glu	Ala	Arg	335	340	345
Met	Asn	Phe	Asp	Lys	Arg	Gln	His	Glu	Ala	Arg	Ile	Gln	Gln	Met	350	355	360
Glu	Asn	Glu	Ile	His	Tyr	Leu	Gln	Glu	Asn	Leu	Lys	Ser	Met	Glu	365	370	375
Glu	Ile	Gln	Gly	Leu	Thr	Asp	Leu	Gln	Leu	Gln	Glu	Ala	Asp	Glu	380	385	390
Glu	Lys	Glu	Arg	Ile	Leu	Ala	Gln	Leu	Arg	Glu	Leu	Glu	Lys	Lys	395	400	405
Lys	Lys	Leu	Glu	Asp	Ala	Lys	Ser	Gln	Glu	Gln	Val	Phe	Gly	Leu	410	415	420
Asp	Lys	Glu	Leu	Lys	Lys	Leu	Lys	Lys	Ala	Val	Ala	Thr	Ser	Asp	425	430	435
Lys	Leu	Ala	Thr	Ala	Glu	Leu	Thr	Ile	Ala	Lys	Asp	Gln	Leu	Lys	440	445	450
Ser	Leu	His	Gly	Thr	Val	Met	Lys	Ile	Asn	Gln	Glu	Arg	Ala	Glu	455	460	465
Glu	Leu	Gln	Glu	Ala	Glu	Arg	Phe	Ser	Arg	Lys	Ala	Ala	Gln	Ala	470	475	480
Ala	Arg	Asp	Leu	Thr	Arg	Ala	Glu	Ala	Glu	Ile	Glu	Leu	Leu	Gln	485	490	495
Asn	Leu	Leu	Arg	Gln	Lys	Gly	Glu	Gln	Phe	Arg	Leu	Glu	Met	Glu	500	505	510
Lys	Thr	Gly	Val	Gly	Thr	Gly	Ala	Asn	Ser	Gln	Val	Leu	Glu	Ile	515	520	525
Glu	Lys	Leu	Asn	Glu	Thr	Met	Glu	Arg	Gln	Arg	Thr	Glu	Ile	Ala	530	535	540
Arg	Leu	Gln	Asn	Val	Leu	Asp	Leu	Thr	Gly	Ser	Asp	Asn	Lys	Gly	545	550	555
Gly	Phe	Glu	Asn	Val	Leu	Glu	Glu	Ile	Ala	Glu	Leu	Arg	Arg	Glu	560	565	570
Val	Ser	Tyr	Gln	Asn	Asp	Tyr	Ile	Ser	Ser	Met	Ala	Asp	Pro	Phe	575	580	585

Lys	Arg	Arg	Gly	Tyr	Trp	Tyr	Phe	Met	Pro	Pro	Pro	Pro	Ser	Ser	590	595	600
Lys	Val	Ser	Ser	His	Ser	Ser	Gln	Ala	Thr	Lys	Asp	Ser	Gly	Val	605	610	615
Gly	Leu	Lys	Tyr	Ser	Ala	Ser	Thr	Pro	Val	Arg	Lys	Pro	Arg	Pro	620	625	630
Gly	Gln	Gln	Asp	Gly	Lys	Glu	Gly	Ser	Gln	Pro	Pro	Pro	Ala	Ser	635	640	645
Gly	Tyr	Trp	Val	Tyr	Ser	Pro	Ile	Arg	Ser	Gly	Leu	His	Lys	Leu	650	655	660
Phe	Pro	Ser	Arg	Asp	Ala	Asp	Ser	Gly	Gly	Asp	Ser	Gln	Glu	Glu	665	670	675
Ser	Glu	Leu	Asp	Asp	Gln	Glu	Glu	Pro	Pro	Phe	Val	Pro	Pro	Pro	680	685	690
Gly	Tyr	Met	Met	Tyr	Thr	Val	Leu	Pro	Asp	Gly	Ser	Pro	Val	Pro	695	700	705
Gln	Gly	Met	Ala	Leu	Tyr	Ala	Pro	Pro	Pro	Pro	Leu	Pro	Asn	Asn	710	715	720
Ser	Arg	Pro	Leu	Thr	Pro	Gly	Thr	Val	Val	Tyr	Gly	Pro	Pro	Pro	725	730	735
Ala	Gly	Ala	Pro	Met	Val	Tyr	Gly	Pro	Pro	Pro	Pro	Asn	Phe	Ser	740	745	750
Ile	Pro	Phe	Ile	Pro	Met	Gly	Val	Leu	His	Cys	Asn	Val	Pro	Glu	755	760	765
His	His	Asn	Leu	Glu	Asn	Glu	Val	Ser	Arg	Leu	Glu	Asp	Ile	Met	770	775	780
Gln	His	Leu	Lys	Ser	Lys	Lys	Arg	Glu	Glu	Arg	Trp	Met	Arg	Ala	785	790	795
Ser	Lys	Arg	Gln	Ser	Glu	Lys	Glu	Met	Glu	Glu	Leu	His	His	Asn	800	805	810
Ile	Asp	Asp	Leu	Leu	Gln	Glu	Lys	Lys	Ser	Leu	Glu	Cys	Glu	Val	815	820	825
Glu	Glu	Leu	His	Arg	Thr	Val	Gln	Lys	Arg	Gln	Gln	Gln	Lys	Asp	830	835	840
Phe	Ile	Asp	Gly	Asn	Val	Glu	Ser	Leu	Met	Thr	Glu	Leu	Glu	Ile	845	850	855
Glu	Lys	Ser	Leu	Lys	His	His	Glu	Asp	Ile	Val	Asp	Glu	Ile	Glu	860	865	870
Cys	Ile	Glu	Lys	Thr	Leu	Leu	Lys	Arg	Arg	Ser	Glu	Leu	Arg	Glu	875	880	885
Ala	Asp	Arg	Leu	Leu	Ala	Glu	Ala	Glu	Ser	Glu	Leu	Ser	Cys	Thr	890	895	900
Lys	Glu	Lys	Thr	Lys	Asn	Ala	Val	Glu	Lys	Phe	Thr	Asp	Ala	Lys	905	910	915
Arg	Ser	Leu	Leu	Gln	Thr	Glu	Ser	Asp	Ala	Glu	Glu	Leu	Glu	Arg	920	925	930
Arg	Ala	Gln	Glu	Thr	Ala	Val	Asn	Leu	Val	Lys	Ala	Asp	Gln	Gln	935	940	945
Leu	Arg	Ser	Leu	Gln	Ala	Asp	Ala	Lys	Asp	Leu	Glu	Gln	His	Lys	950	955	960
Ile	Lys	Gln	Glu	Glu	Ile	Leu	Lys	Glu	Ile	Asn	Lys	Ile	Val	Ala	965	970	975
Ala	Lys	Asp	Ser	Asp	Phe	Gln	Cys	Leu	Ser	Lys	Lys	Lys	Glu	Lys	980	985	990
Leu	Thr	Glu	Glu	Leu	Gln	Lys	Leu	Gln	Lys	Asp	Ile	Glu	Met	Ala	995	1000	1005

Glu Arg Asn Glu Asp His His Leu Gln Val Leu Lys Glu Ser Glu		
1010	1015	1020
Val Leu Leu Gln Ala Lys Arg Ala Glu Leu Glu Lys Leu Lys Ser		
1025	1030	1035
Gln Val Thr Ser Gln Gln Gln Glu Met Ala Val Leu Asp Arg Gln		
1040	1045	1050
Leu Gly His Lys Lys Glu Glu Leu His Leu Leu Gln Gly Ser Met		
1055	1060	1065
Val Gln Ala Lys Ala Asp Leu Gln Glu Ala Leu Arg Leu Gly Glu		
1070	1075	1080
Thr Glu Val Thr Glu Lys Cys Asn His Ile Arg Glu Val Lys Ser		
1085	1090	1095
Leu Leu Glu Glu Leu Ser Phe Gln Lys Gly Glu Leu Asn Val Gln		
1100	1105	1110
Ile Ser Glu Arg Lys Thr Gln Leu Thr Leu Ile Lys Gln Glu Ile		
1115	1120	1125
Glu Lys Glu Glu Glu Asn Leu Gln Val Val Leu Arg Gln Met Ser		
1130	1135	1140
Lys His Lys Thr Glu Leu Lys Asn Ile Leu Asp Met Leu Gln Leu		
1145	1150	1155
Glu Asn His Glu Leu Gln Gly Leu Lys Leu Gln His Asp Gln Arg		
1160	1165	1170
Val Ser Glu Leu Glu Lys Thr Gln Val Ala Val Leu Glu Glu Lys		
1175	1180	1185
Leu Glu Leu Glu Asn Leu Gln Gln Ile Ser Gln Gln Gln Lys Gly		
1190	1195	1200
Glu Ile Glu Trp Gln Lys Gln Leu Leu Glu Arg Asp Lys Arg Glu		
1205	1210	1215
Ile Glu Arg Met Thr Ala Glu Ser Arg Ala Leu Gln Ser Cys Val		
1220	1225	1230
Glu Cys Leu Ser Lys Glu Lys Glu Asp Leu Gln Glu Lys Cys Asp		
1235	1240	1245
Ile Trp Glu Lys Lys Leu Ala Gln Thr Lys Arg Val Leu Ala Ala		
1250	1255	1260
Ala Glu Glu Asn Ser Lys Met Glu Gln Ser Asn Leu Glu Lys Leu		
1265	1270	1275
Glu Leu Asn Val Arg Lys Leu Gln Gln Glu Leu Asp Gln Leu Asn		
1280	1285	1290
Arg Asp Lys Leu Ser Leu His Asn Asp Ile Ser Ala Met Gln Gln		
1295	1300	1305
Gln Leu Gln Glu Lys Arg Glu Ala Val Asn Ser Leu Gln Glu Glu		
1310	1315	1320
Leu Ala Asn Val Gln Asp His Leu Asn Leu Ala Lys Gln Asp Leu		
1325	1330	1335
Leu His Thr Thr Lys His Gln Asp Val Leu Leu Ser Glu Gln Thr		
1340	1345	1350
Arg Leu Gln Lys Asp Ile Ser Glu Trp Ala Asn Arg Phe Glu Asp		
1355	1360	1365
Cys Gln Lys Glu Glu Glu Thr Lys Gln Gln Gln Leu Gln Val Leu		
1370	1375	1380
Gln Asn Glu Ile Glu Glu Asn Lys Leu Lys Leu Val Gln Gln Glu		
1385	1390	1395
Met Met Phe Gln Arg Leu Gln Lys Glu Arg Glu Ser Glu Glu Ser		
1400	1405	1410
Lys Leu Glu Thr Ser Lys Val Thr Leu Lys Glu Gln Gln His Gln		
1415	1420	1425

Leu Glu Lys Glu Leu Thr Asp Gln Lys Ser Lys Leu Asp Gln Val  
 1430 1435 1440  
 Leu Ser Lys Val Leu Ala Ala Glu Glu Arg Val Arg Thr Leu Gln  
 1445 1450 1455  
 Glu Glu Glu Arg Trp Cys Glu Ser Leu Glu Lys Thr Leu Ser Gln  
 1460 1465 1470  
 Thr Lys Arg Gln Leu Ser Glu Arg Glu Gln Gln Leu Val Glu Lys  
 1475 1480 1485  
 Ser Gly Glu Leu Leu Ala Leu Gln Lys Glu Ala Asp Ser Met Arg  
 1490 1495 1500  
 Ala Asp Phe Ser Leu Leu Arg Asn Gln Phe Leu Thr Glu Arg Lys  
 1505 1510 1515  
 Lys Ala Glu Lys Gln Val Ala Ser Leu Lys Glu Ala Leu Lys Ile  
 1520 1525 1530  
 Gln Arg Ser Gln Leu Glu Lys Asn Leu Leu Met Ala Asn Gln Lys  
 1535 1540 1545  
 Asp Leu Glu Arg Arg Gln Met Glu Ile Ser Asp Ala Met Arg Thr  
 1550 1555 1560  
 Leu Lys Ser Glu Val Lys Asp Glu Ile Arg Thr Ser Leu Lys Asn  
 1565 1570 1575  
 Leu Asn Gln Phe Leu Pro Glu Leu Pro Ala Asp Leu Glu Ala Ile  
 1580 1585 1590  
 Leu Glu Arg Asn Glu Asn Leu Glu Gly Glu Leu Glu Ser Leu Lys  
 1595 1600 1605  
 Glu Asn Leu Pro Phe Thr Met Asn Glu Gly Pro Phe Glu Glu Lys  
 1610 1615 1620  
 Leu Asn Phe Ser Gln Val His Ile Met Asp Glu His Trp Arg Gly  
 1625 1630 1635  
 Glu Ala Leu Arg Glu Lys Leu Arg His Arg Glu Asp Arg Leu Lys  
 1640 1645 1650  
 Ala Gln Leu Arg His Cys Met Ser Lys Gln Ala Glu Val Leu Ile  
 1655 1660 1665  
 Lys Gly Lys Arg Gln Thr Glu Gly Thr Leu His Ser Leu Arg Arg  
 1670 1675 1680  
 Gln Val Asp Ala Leu Gly Glu Leu Val Thr Ser Thr Ser Ala Asp  
 1685 1690 1695  
 Ser Ala Ser Ser Pro Ser Leu Ser Gln Leu Glu Ser Ser Leu Thr  
 1700 1705 1710  
 Glu Asp Ser Gln Leu Gly Gln Asn Gln Glu Lys Asn Ala Ser Ala  
 1715 1720 1725  
 Arg

&lt;210&gt; 14

&lt;211&gt; 747

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1482140CD1

&lt;400&gt; 14

Met Ala Leu Gly Leu Gln Arg Ala Arg Pro Ala Leu Ser Cys Gly  
 1 5 10 15  
 Val Ile Ser Pro Pro Cys Ala Pro Thr Arg Asn Ser His Pro Gly

	20		25		30
Pro Gly Cys Thr	Ala Ser Pro Pro Ala	Pro Pro Gly Trp Pro Phe			
	35	40		45	
Ser Gln Arg Gly	Pro Gly Arg Trp Ser Thr	Thr Thr Glu Leu Arg Lys			
	50	55		60	
Glu Lys Ser Arg	Asp Ala Ala Arg Ser Arg	Arg Ser Gln Glu Thr			
	65	70		75	
Glu Val Leu Tyr	Gln Leu Ala His Thr Leu	Pro Phe Ala Arg Gly			
	80	85		90	
Val Ser Ala His	Leu Asp Lys Ala Ser Ile	Met Arg Leu Thr Ile			
	95	100		105	
Ser Tyr Leu Arg	Met His Arg Leu Cys Ala	Gly Glu Trp Asn			
	110	115		120	
Gln Val Gly Ala	Gly Glu Asn His Trp Met	Leu Leu Leu Lys Ala			
	125	130		135	
Leu Glu Gly Phe	Val Met Val Leu Thr Ala	Glu Gly Asp Met Ala			
	140	145		150	
Tyr Leu Ser Glu	Asn Val Ser Lys His Leu	Gly Leu Ser Gln Leu			
	155	160		165	
Glu Leu Ile Gly	His Ser Ile Phe Asp Phe	Ile His Pro Cys Asp			
	170	175		180	
Gln Glu Glu Leu	Gln Asp Ala Leu Thr Pro	Gln Gln Thr Leu Ser			
	185	190		195	
Arg Arg Lys Val	Glu Ala Pro Thr Glu Arg	Cys Phe Ser Leu Arg			
	200	205		210	
Met Lys Ser Thr	Leu Thr Ser Arg Gly Arg	Thr Leu Asn Leu Lys			
	215	220		225	
Ala Ala Thr Trp	Lys Val Leu Asn Cys Ser	Gly His Met Arg Ala			
	230	235		240	
Tyr Lys Pro Pro	Ala Gln Thr Ser Pro Ala	Gly Ser Pro Asp Ser			
	245	250		255	
Glu Pro Pro Leu	Gln Cys Leu Val Leu Ile	Cys Glu Ala Ile Pro			
	260	265		270	
His Pro Gly Ser	Leu Glu Pro Pro Leu Gly	Arg Gly Ala Phe Leu			
	275	280		285	
Ser Arg His Ser	Leu Asp Met Lys Phe Thr	Tyr Cys Asp Asp Arg			
	290	295		300	
Ile Ala Glu Val	Ala Gly Tyr Ser Pro Asp	Asp Leu Ile Gly Cys			
	305	310		315	
Ser Ala Tyr Glu	Tyr Ile His Ala Leu Asp	Ser Asp Ala Val Ser			
	320	325		330	
Lys Ser Ile His	Thr Leu Leu Ser Lys Gly	Gln Ala Val Thr Gly			
	335	340		345	
Gln Tyr Arg Phe	Leu Ala Arg Ser Gly Gly	Tyr Leu Trp Thr Gln			
	350	355		360	
Thr Gln Ala Thr	Val Val Ser Gly Gly Arg	Gly Pro Gln Ser Glu			
	365	370		375	
Ser Ile Val Cys	Val His Phe Leu Ile Ser	Gln Val Glu Glu Thr			
	380	385		390	
Gly Val Val Leu	Ser Leu Glu Gln Thr Glu	Gln His Ser Arg Arg			
	395	400		405	
Pro Ile Gln Arg	Gly Ala Pro Ser Gln Lys	Asp Thr Pro Asn Pro			
	410	415		420	
Gly Asp Ser Leu	Asp Thr Pro Gly Pro Arg	Ile Leu Ala Phe Leu			
	425	430		435	
His Pro Pro Ser	Leu Ser Glu Ala Ala Leu	Ala Ala Asp Pro Arg			

	440		445		450
Arg Phe Cys Ser Pro Asp Leu Arg Arg	Leu Leu Gly Pro Ile Leu				
	455		460		465
Asp Gly Ala Ser Val Ala Ala Thr Pro	Ser Thr Pro Leu Ala Thr				
	470		475		480
Arg His Pro Gln Ser Pro Leu Ser Ala	Asp Leu Pro Asp Glu Leu				
	485		490		495
Pro Val Gly Thr Glu Asn Val His Arg	Leu Phe Thr Ser Gly Lys				
	500		505		510
Asp Thr Glu Ala Val Glu Thr Asp Leu	Asp Ile Ala Gln Asp Ala				
	515		520		525
Asp Ala Leu Asp Leu Glu Met Leu Ala	Pro Tyr Ile Ser Met Asp				
	530		535		540
Asp Asp Phe Gln Leu Asn Ala Ser Glu	Gln Leu Pro Arg Ala Tyr				
	545		550		555
His Arg Pro Leu Gly Ala Val Pro Arg	Pro Arg Ala Arg Ser Phe				
	560		565		570
His Gly Leu Ser Pro Pro Ala Leu Glu	Pro Ser Leu Leu Pro Arg				
	575		580		585
Trp Gly Ser Asp Pro Arg Leu Ser Cys	Ser Ser Pro Ser Arg Gly				
	590		595		600
Asp Pro Ser Ala Ser Ser Pro Met Ala	Gly Ala Arg Lys Arg Thr				
	605		610		615
Leu Ala Gln Ser Ser Glu Asp Glu Asp	Glu Gly Val Glu Leu Leu				
	620		625		630
Gly Val Arg Pro Pro Lys Arg Ser Pro	Ser Pro Glu His Glu Asn				
	635		640		645
Phe Leu Leu Phe Pro Leu Ser Leu Ser	Phe Leu Leu Thr Gly Gly				
	650		655		660
Pro Ala Pro Gly Ser Leu Gln Asp Pro	Thr Glu Leu Thr Gln Phe				
	665		670		675
Leu Leu Ser Val Leu Ser Phe Pro Ile	Leu Asp Pro Tyr Pro Leu				
	680		685		690
Gly Cys Ala Ala Pro Gly Leu His Ala	Ser Pro Phe Ser Leu Pro				
	695		700		705
Thr Ile Ser Val Pro Gln Asn Pro Leu	His Ser Pro Pro Gln Pro				
	710		715		720
Ser Arg His Ala Leu Thr Leu Thr Leu	Pro His Met Phe Gly Ala				
	725		730		735
Pro Gly Ala Pro Ser Pro Leu Gly Trp	Phe Ala Ile				
	740		745		

&lt;210&gt; 15

&lt;211&gt; 2759

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 394992CD1

&lt;400&gt; 15

Met Val Ser Glu Glu Glu Glu Glu Glu Asp Gly Asp Ala Glu Glu

1

5

10

15

Thr Gln Asp Ser Glu Asp Asp Glu Glu Asp Glu Met Glu Glu Asp

20

25

30

Asp	Asp	Asp	Ser	Asp	Tyr	Pro	Glu	Glu	Met	Glu	Asp	Asp	Asp	Asp
				35					40					45
Asp	Ala	Ser	Tyr	Cys	Thr	Glu	Ser	Ser	Phe	Arg	Ser	His	Ser	Thr
				50					55					60
Tyr	Ser	Ser	Thr	Pro	Gly	Arg	Arg	Lys	Pro	Arg	Val	His	Arg	Pro
				65					70					75
Arg	Ser	Pro	Ile	Leu	Glu	Glu	Lys	Asp	Ile	Pro	Pro	Leu	Glu	Phe
				80					85					90
Pro	Lys	Ser	Ser	Glu	Asp	Leu	Met	Val	Pro	Asn	Glu	His	Ile	Met
				95					100					105
Asn	Val	Ile	Ala	Ile	Tyr	Glu	Val	Leu	Arg	Asn	Phe	Gly	Thr	Val
				110					115					120
Leu	Arg	Leu	Ser	Pro	Phe	Arg	Phe	Glu	Asp	Phe	Cys	Ala	Ala	Leu
				125					130					135
Val	Ser	Gln	Glu	Gln	Cys	Thr	Leu	Met	Ala	Glu	Met	His	Val	Val
				140					145					150
Leu	Leu	Lys	Ala	Val	Leu	Arg	Glu	Glu	Asp	Thr	Ser	Asn	Thr	Thr
				155					160					165
Phe	Gly	Pro	Ala	Asp	Leu	Lys	Asp	Ser	Val	Asn	Ser	Thr	Leu	Tyr
				170					175					180
Phe	Ile	Asp	Gly	Met	Thr	Trp	Pro	Glu	Val	Leu	Arg	Val	Tyr	Cys
				185					190					195
Glu	Ser	Asp	Lys	Glu	Tyr	His	His	Val	Leu	Pro	Tyr	Gln	Glu	Ala
				200					205					210
Glu	Asp	Tyr	Pro	Tyr	Gly	Pro	Val	Glu	Asn	Lys	Ile	Lys	Val	Leu
				215					220					225
Gln	Phe	Leu	Val	Asp	Gln	Phe	Leu	Thr	Thr	Asn	Ile	Ala	Arg	Glu
				230					235					240
Glu	Leu	Met	Ser	Glu	Gly	Val	Ile	Gln	Tyr	Asp	Asp	His	Cys	Arg
				245					250					255
Val	Cys	His	Lys	Leu	Gly	Asp	Leu	Leu	Cys	Cys	Glu	Thr	Cys	Ser
				260					265					270
Ala	Val	Tyr	His	Leu	Glu	Cys	Val	Lys	Pro	Pro	Leu	Glu	Glu	Val
				275					280					285
Pro	Glu	Asp	Glu	Trp	Gln	Cys	Glu	Val	Cys	Val	Ala	His	Lys	Val
				290					295					300
Pro	Gly	Val	Thr	Asp	Cys	Val	Ala	Glu	Ile	Gln	Lys	Asn	Lys	Pro
				305					310					315
Tyr	Ile	Arg	His	Glu	Pro	Ile	Gly	Tyr	Asp	Arg	Ser	Arg	Arg	Lys
				320					325					330
Tyr	Trp	Phe	Leu	Asn	Arg	Arg	Leu	Ile	Ile	Glu	Glu	Asp	Thr	Glu
				335					340					345
Asn	Glu	Asn	Glu	Lys	Lys	Ile	Trp	Tyr	Tyr	Ser	Thr	Lys	Val	Gln
				350					355					360
Leu	Ala	Glu	Leu	Ile	Asp	Cys	Leu	Asp	Lys	Asp	Tyr	Trp	Glu	Ala
				365					370					375
Glu	Leu	Cys	Lys	Ile	Leu	Glu	Glu	Met	Arg	Glu	Glu	Ile	His	Arg
				380					385					390
His	Met	Asp	Ile	Thr	Glu	Asp	Leu	Thr	Asn	Lys	Ala	Arg	Gly	Ser
				395					400					405
Asn	Lys	Ser	Phe	Leu	Ala	Ala	Ala	Asn	Glu	Glu	Ile	Leu	Glu	Ser
				410					415					420
Ile	Arg	Ala	Lys	Lys	Gly	Asp	Ile	Asp	Asn	Val	Lys	Ser	Pro	Glu
				425					430					435
Glu	Thr	Glu	Lys	Asp	Lys	Asn	Glu	Thr	Glu	Asn	Asp	Ser	Lys	Asp
				440					445					450

Ala Glu Lys Asn Arg Glu Glu Phe Glu Asp Gln Ser Leu Glu Lys	455	460	465
Asp Ser Asp Asp Lys Thr Pro Asp Asp Asp Pro Glu Gln Gly Lys	470	475	480
Ser Glu Glu Pro Thr Glu Val Gly Asp Lys Gly Asn Ser Val Ser	485	490	495
Ala Asn Leu Gly Asp Asn Thr Thr Asn Ala Thr Ser Glu Glu Thr	500	505	510
Ser Pro Ser Glu Gly Arg Ser Pro Val Gly Cys Leu Ser Glu Thr	515	520	525
Pro Asp Ser Ser Asn Met Ala Glu Lys Lys Val Ala Ser Glu Leu	530	535	540
Pro Gln Asp Val Pro Val Gly Asp Phe Lys Ser Glu Lys Ser Asn	545	550	555
Gly Glu Leu Ser Glu Ser Pro Gly Ala Gly Lys Gly Ala Ser Gly	560	565	570
Ser Thr Arg Ile Ile Thr Arg Leu Arg Asn Pro Asp Ser Lys Leu	575	580	585
Ser Gln Leu Lys Ser Gln Gln Val Ala Ala Ala Ala His Glu Ala	590	595	600
Asn Lys Leu Phe Lys Glu Gly Lys Glu Val Leu Val Val Asn Ser	605	610	615
Gln Gly Glu Ile Ser Arg Leu Ser Thr Lys Lys Glu Val Ile Met	620	625	630
Lys Gly Asn Ile Asn Asn Tyr Phe Lys Leu Gly Gln Glu Gly Lys	635	640	645
Tyr Arg Val Tyr His Asn Gln Tyr Ser Thr Asn Ser Phe Ala Leu	650	655	660
Asn Lys His Gln His Arg Glu Asp His Asp Lys Arg Arg His Leu	665	670	675
Ala His Lys Phe Cys Leu Thr Pro Ala Gly Glu Phe Lys Trp Asn	680	685	690
Gly Ser Val His Gly Ser Lys Val Leu Thr Ile Ser Thr Leu Arg	695	700	705
Leu Thr Ile Thr Gln Leu Glu Asn Asn Ile Pro Ser Ser Phe Leu	710	715	720
His Pro Asn Trp Ala Ser His Arg Ala Asn Trp Ile Lys Ala Val	725	730	735
Gln Met Cys Ser Lys Pro Arg Glu Phe Ala Leu Ala Leu Ala Ile	740	745	750
Leu Glu Cys Ala Val Lys Pro Val Val Met Leu Pro Ile Trp Arg	755	760	765
Glu Ser Leu Gly His Thr Arg Leu His Arg Met Thr Ser Ile Glu	770	775	780
Arg Glu Glu Lys Glu Lys Val Lys Lys Lys Glu Lys Lys Gln Glu	785	790	795
Glu Glu Glu Thr Met Gln Gln Ala Thr Trp Val Lys Tyr Thr Phe	800	805	810
Pro Val Lys His Gln Val Trp Lys Gln Lys Gly Glu Glu Tyr Arg	815	820	825
Val Thr Gly Tyr Gly Gly Trp Ser Trp Ile Ser Lys Thr His Val	830	835	840
Tyr Arg Phe Val Pro Lys Leu Pro Gly Asn Thr Asn Val Asn Tyr	845	850	855
Arg Lys Ser Leu Glu Gly Thr Lys Asn Asn Met Asp Glu Asn Met	860	865	870

Asp	Glu	Ser	Asp	Lys	Arg	Lys	Cys	Ser	Arg	Ser	Pro	Lys	Lys	Ile	875	880	885
Lys	Ile	Glu	Pro	Asp	Ser	Glu	Lys	Asp	Glu	Val	Lys	Gly	Ser	Asp	890	895	900
Ala	Ala	Lys	Gly	Ala	Asp	Gln	Asn	Glu	Met	Asp	Ile	Ser	Lys	Ile	905	910	915
Thr	Glu	Lys	Lys	Asp	Gln	Asp	Val	Lys	Glu	Leu	Leu	Asp	Ser	Asp	920	925	930
Ser	Asp	Lys	Pro	Cys	Lys	Glu	Glu	Pro	Met	Glu	Val	Asp	Asp	Asp	935	940	945
Met	Lys	Thr	Glu	Ser	His	Val	Asn	Cys	Gln	Glu	Ser	Ser	Gln	Val	950	955	960
Asp	Val	Val	Asn	Val	Ser	Glu	Gly	Phe	His	Leu	Arg	Thr	Ser	Tyr	965	970	975
Lys	Lys	Lys	Thr	Lys	Ser	Ser	Lys	Leu	Asp	Gly	Leu	Leu	Glu	Arg	980	985	990
Arg	Ile	Lys	Gln	Phe	Thr	Leu	Glu	Glu	Lys	Gln	Arg	Leu	Glu	Lys	995	1000	1005
Ile	Lys	Leu	Glu	Gly	Gly	Ile	Lys	Gly	Ile	Gly	Lys	Thr	Ser	Thr	1010	1015	1020
Asn	Ser	Ser	Lys	Asn	Leu	Ser	Glu	Ser	Pro	Val	Ile	Thr	Lys	Ala	1025	1030	1035
Lys	Glu	Gly	Cys	Gln	Ser	Asp	Ser	Met	Arg	Gln	Glu	Gln	Ser	Pro	1040	1045	1050
Asn	Ala	Asn	Asn	Asp	Gln	Pro	Glu	Asp	Leu	Ile	Gln	Gly	Cys	Ser	1055	1060	1065
Glu	Ser	Asp	Ser	Ser	Val	Leu	Arg	Met	Ser	Asp	Pro	Ser	His	Thr	1070	1075	1080
Thr	Asn	Lys	Leu	Tyr	Pro	Lys	Asp	Arg	Val	Leu	Asp	Asp	Val	Ser	1085	1090	1095
Ile	Arg	Ser	Pro	Glu	Thr	Lys	Cys	Pro	Lys	Gln	Asn	Ser	Ile	Glu	1100	1105	1110
Asn	Asp	Ile	Glu	Glu	Lys	Val	Ser	Asp	Leu	Ala	Ser	Arg	Gly	Gln	1115	1120	1125
Glu	Pro	Ser	Lys	Ser	Lys	Thr	Lys	Gly	Asn	Asp	Phe	Phe	Ile	Asp	1130	1135	1140
Asp	Ser	Lys	Leu	Ala	Ser	Ala	Asp	Asp	Ile	Gly	Thr	Leu	Ile	Cys	1145	1150	1155
Lys	Asn	Lys	Lys	Pro	Leu	Ile	Gln	Glu	Glu	Ser	Asp	Thr	Ile	Val	1160	1165	1170
Ser	Ser	Ser	Lys	Ser	Ala	Leu	His	Ser	Ser	Val	Pro	Lys	Ser	Thr	1175	1180	1185
Asn	Asp	Arg	Asp	Ala	Thr	Pro	Leu	Ser	Arg	Ala	Met	Asp	Phe	Glu	1190	1195	1200
Gly	Lys	Leu	Gly	Cys	Asp	Ser	Glu	Ser	Asn	Ser	Thr	Leu	Glu	Asn	1205	1210	1215
Ser	Ser	Asp	Thr	Val	Ser	Ile	Gln	Asp	Ser	Ser	Glu	Glu	Asp	Met	1220	1225	1230
Ile	Val	Gln	Asn	Ser	Asn	Glu	Ser	Ile	Ser	Glu	Gln	Phe	Arg	Thr	1235	1240	1245
Arg	Glu	Gln	Asp	Val	Glu	Val	Leu	Glu	Pro	Leu	Lys	Cys	Glu	Leu	1250	1255	1260
Val	Ser	Gly	Glu	Ser	Thr	Gly	Asn	Cys	Glu	Asp	Arg	Leu	Pro	Val	1265	1270	1275
Lys	Gly	Thr	Glu	Ala	Asn	Gly	Lys	Lys	Pro	Ser	Gln	Gln	Lys	Lys	1280	1285	1290

Leu Glu Glu Arg Pro Val Asn Lys Cys Ser Asp Gln Ile Lys Leu	1295	1300	1305
Lys Asn Thr Thr Asp Lys Lys Asn Asn Glu Asn Arg Glu Ser Glu	1310	1315	1320
Lys Lys Gly Gln Arg Thr Ser Thr Phe Gln Ile Asn Gly Lys Asp	1325	1330	1335
Asn Lys Pro Lys Ile Tyr Leu Lys Gly Glu Cys Leu Lys Glu Ile	1340	1345	1350
Ser Glu Ser Arg Val Val Ser Gly Asn Val Glu Pro Lys Val Asn	1355	1360	1365
Asn Ile Asn Lys Ile Ile Pro Glu Asn Asp Ile Lys Ser Leu Thr	1370	1375	1380
Val Lys Glu Ser Ala Ile Arg Pro Phe Ile Asn Gly Asp Val Ile	1385	1390	1395
Met Glu Asp Phe Asn Glu Arg Asn Ser Ser Glu Thr Lys Ser His	1400	1405	1410
Leu Leu Ser Ser Ser Asp Ala Glu Gly Asn Tyr Arg Asp Ser Leu	1415	1420	1425
Glu Thr Leu Pro Ser Thr Lys Glu Ser Asp Ser Thr Gln Thr Thr	1430	1435	1440
Thr Pro Ser Ala Ser Cys Pro Glu Ser Asn Ser Val Asn Gln Val	1445	1450	1455
Glu Asp Met Glu Ile Glu Thr Ser Glu Val Lys Lys Val Thr Ser	1460	1465	1470
Ser Pro Ile Thr Ser Glu Glu Glu Ser Asn Leu Ser Asn Asp Phe	1475	1480	1485
Ile Asp Glu Asn Gly Leu Pro Ile Asn Lys Asn Glu Asn Val Asn	1490	1495	1500
Gly Glu Ser Lys Arg Lys Thr Val Ile Thr Glu Val Thr Thr Met	1505	1510	1515
Thr Ser Thr Val Ala Thr Glu Ser Lys Thr Val Ile Lys Val Glu	1520	1525	1530
Lys Gly Asp Lys Gln Thr Val Val Ser Ser Thr Glu Asn Cys Ala	1535	1540	1545
Lys Ser Thr Val Thr Thr Thr Thr Thr Val Thr Lys Leu Ser	1550	1555	1560
Thr Pro Ser Thr Gly Gly Ser Val Asp Ile Ile Ser Val Lys Glu	1565	1570	1575
Gln Ser Lys Thr Val Val Thr Thr Thr Val Thr Asp Ser Leu Thr	1580	1585	1590
Thr Thr Gly Gly Thr Leu Val Thr Ser Met Thr Val Ser Lys Glu	1595	1600	1605
Tyr Ser Thr Arg Asp Lys Val Lys Leu Met Lys Phe Ser Arg Pro	1610	1615	1620
Lys Lys Thr Arg Ser Gly Thr Ala Leu Pro Ser Tyr Arg Lys Phe	1625	1630	1635
Val Thr Lys Ser Ser Lys Lys Ser Ile Phe Val Leu Pro Asn Asp	1640	1645	1650
Asp Leu Lys Lys Leu Ala Arg Lys Gly Gly Ile Arg Glu Val Pro	1655	1660	1665
Tyr Phe Asn Tyr Asn Ala Lys Pro Ala Leu Asp Ile Trp Pro Tyr	1670	1675	1680
Pro Ser Pro Arg Pro Thr Phe Gly Ile Thr Trp Arg Tyr Arg Leu	1685	1690	1695
Gln Thr Val Lys Ser Leu Ala Gly Val Ser Leu Met Leu Arg Leu	1700	1705	1710

Leu Trp Ala Ser Leu Arg Trp Asp Asp Met Ala Ala Lys Ala Pro	1715	1720	1725
Pro Gly Gly Gly Thr Thr Arg Thr Glu Thr Ser Glu Thr Glu Ile	1730	1735	1740
Thr Thr Thr Glu Ile Ile Lys Arg Arg Asp Val Gly Pro Tyr Gly	1745	1750	1755
Ile Arg Ser Glu Tyr Cys Ile Arg Lys Ile Ile Cys Pro Ile Gly	1760	1765	1770
Val Pro Glu Thr Pro Lys Glu Thr Pro Thr Pro Gln Arg Lys Gly	1775	1780	1785
Leu Arg Ser Ser Ala Leu Arg Pro Lys Arg Pro Glu Thr Pro Lys	1790	1795	1800
Gln Thr Gly Pro Val Ile Ile Glu Thr Trp Val Ala Glu Glu Glu	1805	1810	1815
Leu Glu Leu Trp Glu Ile Arg Ala Phe Ala Glu Arg Val Glu Lys	1820	1825	1830
Glu Lys Ala Gln Ala Val Glu Gln Gln Ala Lys Lys Arg Leu Glu	1835	1840	1845
Gln Gln Lys Pro Thr Val Ile Ala Thr Ser Thr Thr Ser Pro Thr	1850	1855	1860
Ser Ser Thr Thr Ser Thr Ile Ser Pro Ala Gln Lys Val Met Val	1865	1870	1875
Ala Pro Ile Ser Gly Ser Val Thr Thr Gly Thr Lys Met Val Leu	1880	1885	1890
Thr Thr Lys Val Gly Ser Pro Ala Thr Val Thr Phe Gln Gln Asn	1895	1900	1905
Lys Asn Phe His Gln Thr Phe Ala Thr Trp Val Lys Gln Gly Gln	1910	1915	1920
Ser Asn Ser Gly Val Val Gln Val Gln Gln Lys Val Leu Gly Ile	1925	1930	1935
Ile Pro Ser Ser Thr Gly Thr Ser Gln Gln Thr Phe Thr Ser Phe	1940	1945	1950
Gln Pro Arg Thr Ala Thr Val Thr Ile Arg Pro Asn Thr Ser Gly	1955	1960	1965
Ser Gly Gly Thr Thr Ser Asn Ser Gln Val Ile Thr Gly Pro Gln	1970	1975	1980
Ile Arg Pro Gly Met Thr Val Ile Arg Thr Pro Leu Gln Gln Ser	1985	1990	1995
Thr Leu Gly Lys Ala Ile Ile Arg Thr Pro Val Met Val Gln Pro	2000	2005	2010
Gly Ala Pro Gln Gln Val Met Thr Gln Ile Ile Arg Gly Gln Pro	2015	2020	2025
Val Ser Thr Ala Val Ser Ala Pro Asn Thr Val Ser Ser Thr Pro	2030	2035	2040
Gly Gln Lys Ser Leu Thr Ser Ala Thr Ser Thr Ser Asn Ile Gln	2045	2050	2055
Ser Ser Ala Ser Gln Pro Pro Arg Pro Gln Gln Gly Gln Val Lys	2060	2065	2070
Leu Thr Met Ala Gln Leu Thr Gln Leu Thr Gln Gly His Gly Gly	2075	2080	2085
Asn Gln Gly Leu Thr Val Val Ile Gln Gly Gln Gly Gln Thr Thr	2090	2095	2100
Gly Gln Leu Gln Leu Ile Pro Gln Gly Val Thr Val Leu Pro Gly	2105	2110	2115
Pro Gly Gln Gln Leu Met Gln Ala Ala Met Pro Asn Gly Thr Val	2120	2125	2130

Gln Arg Phe Leu Phe Thr Pro Leu Ala Thr Thr Ala Thr Thr Ala	2135	2140	2145
Ser Thr Thr Thr Thr Thr Val Ser Thr Thr Ala Ala Gly Thr Gly	2150	2155	2160
Glu Gln Arg Gln Ser Lys Leu Ser Pro Gln Met Gln Val His Gln	2165	2170	2175
Asp Lys Thr Leu Pro Pro Ala Gln Ser Ser Ser Val Gly Pro Ala	2180	2185	2190
Glu Ala Gln Pro Gln Thr Ala Gln Pro Ser Ala Gln Pro Gln Pro	2195	2200	2205
Gln Thr Gln Pro Gln Ser Pro Ala Gln Pro Glu Val Gln Thr Gln	2210	2215	2220
Pro Glu Val Gln Thr Gln Thr Thr Val Ser Ser His Val Pro Ser	2225	2230	2235
Glu Ala Gln Pro Thr His Ala Gln Ser Ser Lys Pro Gln Val Ala	2240	2245	2250
Ala Gln Ser Gln Pro Gln Ser Asn Val Gln Gly Gln Ser Pro Val	2255	2260	2265
Arg Val Gln Ser Pro Ser Gln Thr Arg Ile Arg Pro Ser Thr Pro	2270	2275	2280
Ser Gln Leu Ser Pro Gly Gln Gln Ser Gln Val Gln Thr Thr Thr	2285	2290	2295
Ser Gln Pro Ile Pro Ile Gln Pro His Thr Ser Leu Gln Ile Pro	2300	2305	2310
Ser Gln Gly Gln Pro Gln Ser Gln Pro Gln Val Val Met Lys His	2315	2320	2325
Asn Ala Val Ile Glu His Leu Lys Gln Lys Lys Ser Met Thr Pro	2330	2335	2340
Ala Glu Arg Glu Glu Asn Gln Arg Met Ile Val Cys Asn Gln Val	2345	2350	2355
Met Lys Tyr Ile Leu Asp Lys Ile Asp Lys Glu Glu Lys Gln Ala	2360	2365	2370
Ala Lys Lys Arg Lys Arg Glu Glu Ser Val Glu Gln Lys Arg Ser	2375	2380	2385
Lys Gln Asn Ala Thr Lys Leu Ser Ala Leu Leu Phe Lys His Lys	2390	2395	2400
Glu Gln Leu Arg Ala Glu Ile Leu Lys Lys Arg Ala Leu Leu Asp	2405	2410	2415
Lys Asp Leu Gln Ile Glu Val Gln Glu Glu Leu Lys Arg Asp Leu	2420	2425	2430
Lys Ile Lys Lys Glu Lys Asp Leu Met Gln Leu Ala Gln Ala Thr	2435	2440	2445
Ala Val Ala Ala Pro Cys Pro Pro Val Thr Pro Ala Pro Pro Ala	2450	2455	2460
Pro Pro Ala Pro Pro Pro Ser Pro Pro Pro Pro Ala Val Gln	2465	2470	2475
His Thr Gly Leu Leu Ser Thr Pro Thr Leu Pro Ala Ala Ser Gln	2480	2485	2490
Lys Arg Lys Arg Glu Glu Lys Asp Ser Ser Ser Lys Ser Lys	2495	2500	2505
Lys Lys Lys Met Ile Ser Thr Thr Ser Lys Glu Thr Lys Lys Asp	2510	2515	2520
Thr Lys Leu Tyr Cys Ile Cys Lys Thr Pro Tyr Asp Glu Ser Lys	2525	2530	2535
Phe Tyr Ile Gly Cys Asp Leu Cys Thr Asn Trp Tyr His Gly Glu	2540	2545	2550

Cys Val Gly Ile Thr Glu Lys Glu Ala Lys Lys Met Asp Val Tyr	2555	2560	2565
Ile Cys Asn Asp Cys Lys Arg Ala Gln Glu Gly Ser Ser Glu Glu	2570	2575	2580
Leu Tyr Cys Ile Cys Arg Thr Pro Tyr Asp Glu Ser Gln Phe Tyr	2585	2590	2595
Ile Gly Cys Asp Arg Cys Gln Asn Trp Tyr His Gly Arg Cys Val	2600	2605	2610
Gly Ile Leu Gln Ser Glu Ala Glu Leu Ile Asp Glu Tyr Val Cys	2615	2620	2625
Pro Gln Cys Gln Ser Thr Glu Asp Ala Met Thr Val Leu Thr Pro	2630	2635	2640
Leu Thr Glu Lys Asp Tyr Glu Gly Leu Lys Arg Val Leu Arg Ser	2645	2650	2655
Leu Gln Ala His Lys Met Ala Trp Pro Phe Leu Glu Pro Val Asp	2660	2665	2670
Pro Asn Asp Ala Pro Asp Tyr Tyr Gly Val Ile Lys Glu Pro Met	2675	2680	2685
Asp Leu Ala Thr Met Glu Glu Arg Val Gln Arg Arg Tyr Tyr Glu	2690	2695	2700
Lys Leu Thr Glu Phe Val Ala Asp Met Thr Lys Ile Phe Asp Asn	2705	2710	2715
Cys Arg Tyr Tyr Asn Pro Ser Asp Ser Pro Phe Tyr Gln Cys Ala	2720	2725	2730
Glu Val Leu Glu Ser Phe Phe Val Gln Lys Leu Lys Gly Phe Lys	2735	2740	2745
Ala Ser Arg Ser His Asn Asn Lys Leu Gln Ser Thr Ala Ser	2750	2755	

&lt;210&gt; 16

&lt;211&gt; 613

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5093550CD1

&lt;400&gt; 16

Met Asp Val Ala Ile Glu Phe Ser Val Glu Glu Trp Gln Cys Leu	1	5	10	15
Asp Thr Ala Gln Gln Asn Leu Tyr Arg Asn Val Met Leu Glu Asn	20	25	30	
Tyr Arg Asn Leu Val Phe Leu Gly Ile Ala Val Ser Lys Pro Asp	35	40	45	
Leu Ile Thr Cys Leu Glu Gln Gly Lys Glu Pro Trp Asn Met Glu	50	55	60	
Arg His Glu Met Val Ala Lys Pro Pro Gly Met Cys Cys Tyr Phe	65	70	75	
Ala Gln Asp Leu Arg Pro Glu Gln Ser Ile Lys Ala Ser Leu Gln	80	85	90	
Arg Ile Ile Leu Arg Lys Tyr Glu Lys Cys Gly His His Asn Leu	95	100	105	
Gln Leu Lys Lys Gly Tyr Lys Ser Val Asp Glu Tyr Lys Val His	110	115	120	
Lys Gly Ser Tyr Asn Gly Phe Asn Gln Cys Leu Thr Thr Thr Gln				

	125		130		135
Ser Lys Ile Phe	Gln Cys Asp Lys Tyr	Val Lys Asp Phe His	Lys		
	140		145		150
Phe Ser Asn Ser	Asn Arg His Lys Thr	Glu Lys Asn Pro Phe	Lys		
	155		160		165
Cys Lys Glu Cys	Gly Lys Ser Phe Cys	Val Leu Ser His Leu	Thr		
	170		175		180
Gln His Lys Arg	Ile His Thr Thr Val	Asn Ser Tyr Lys Leu	Glu		
	185		190		195
Glu Cys Gly Lys	Ala Phe Asn Val Ser	Ser Thr Leu Ser Gln	His		
	200		205		210
Lys Arg Ile His	Thr Gly Gln Lys His	Tyr Lys Cys Glu Glu	Cys		
	215		220		225
Gly Ile Ala Phe	Asn Lys Ser Ser His	Leu Asn Thr His Lys	Ile		
	230		235		240
Ile His Thr Gly	Glu Lys Ser Tyr Lys	Arg Glu Glu Cys Gly	Lys		
	245		250		255
Ala Phe Asn Ile	Ser Ser His Leu Thr	Thr His Lys Ile Ile	His		
	260		265		270
Thr Gly Glu Asn	Ala Tyr Lys Cys Lys	Glu Cys Gly Lys Ala	Phe		
	275		280		285
Asn Gln Ser Ser	Thr Leu Thr Arg His	Lys Ile Ile His Ala	Gly		
	290		295		300
Glu Lys Pro Tyr	Ile Cys Glu His Cys	Gly Arg Ala Phe Asn	Gln		
	305		310		315
Ser Ser Asn Leu	Thr Lys His Lys Arg	Ile His Thr Gly Asp	Lys		
	320		325		330
Phe Tyr Lys Cys	Glu Glu Cys Gly Lys	Ala Phe Asn Val Ser	Ser		
	335		340		345
Thr Leu Thr Gln	His Lys Arg Ile His	Thr Gly Glu Lys Pro	Tyr		
	350		355		360
Lys Cys Glu Glu	Cys Gly Lys Ala Phe	Asn Val Ser Ser Thr	Leu		
	365		370		375
Thr Gln His Lys	Arg Ile His Thr Gly	Glu Lys Pro Tyr Lys	Cys		
	380		385		390
Glu Glu Cys Gly	Lys Ala Phe Asn Thr	Ser Ser His Leu Thr	Thr		
	395		400		405
His Lys Arg Ile	His Thr Gly Glu Lys	Pro Tyr Lys Cys Glu	Glu		
	410		415		420
Cys Gly Lys Ala	Phe Asn Gln Phe Ser	Gln Leu Thr Thr His	Lys		
	425		430		435
Ile Ile His Thr	Gly Glu Lys Pro Tyr	Lys Cys Lys Glu Cys	Gly		
	440		445		450
Lys Ala Phe Lys	Arg Ser Ser Asn Leu	Thr Glu His Arg Ile	Ile		
	455		460		465
His Thr Gly Glu	Lys Pro Tyr Lys Cys	Glu Glu Cys Gly Lys	Ala		
	470		475		480
Phe Asn Leu Ser	Ser His Leu Thr Thr	His Lys Lys Ile His	Thr		
	485		490		495
Gly Glu Lys Pro	Tyr Lys Cys Lys Glu	Cys Gly Lys Ala Phe	Asn		
	500		505		510
Gln Ser Ser Thr	Leu Ala Arg His Lys	Ile Ile His Ala Gly	Glu		
	515		520		525
Lys Pro Tyr Lys	Cys Glu Glu Cys Gly	Lys Ala Phe Tyr Gln	Tyr		
	530		535		540
Ser Asn Leu Thr	Gln His Lys Ile Ile	His Thr Gly Glu Lys	Pro		

	545		550		555
Tyr Lys Cys Glu	Glu Cys Gly Lys Ala	Phe Asn Trp Ser Ser	Thr		
	560		565		570
Leu Thr Lys His	Lys Val Ile His Thr	Gly Glu Lys Pro Tyr	Lys		
	575		580		585
Cys Lys Glu Cys	Gly Lys Ala Phe Asn	Gln Cys Ser Asn Leu	Thr		
	590		595		600
Thr His Lys Lys	Ile His Ala Val Glu	Lys Ser Asp Lys			
	605		610		

&lt;210&gt; 17

&lt;211&gt; 240

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7487977CD1

&lt;400&gt; 17

Met Ser Lys Pro Val Asp His Val Lys Arg Pro Met Asn Ala Phe		
1	5	10 15
Met Val Trp Ser Arg Ala Gln Arg Arg Lys Met Ala Gln Glu Asn		
	20	25 30
Pro Lys Met His Asn Ser Glu Ile Ser Lys Arg Leu Gly Ala Glu		
	35	40 45
Trp Lys Leu Leu Ser Glu Ala Glu Lys Arg Pro Tyr Ile Asp Glu		
	50	55 60
Ala Lys Arg Leu Arg Ala Gln His Met Lys Glu His Pro Asp Tyr		
	65	70 75
Lys Tyr Arg Pro Arg Arg Lys Pro Lys Asn Leu Leu Lys Lys Asp		
	80	85 90
Arg Tyr Val Phe Pro Leu Pro Tyr Leu Gly Asp Thr Asp Pro Leu		
	95	100 105
Lys Ala Ala Gly Leu Pro Val Gly Ala Ser Asp Gly Leu Leu Ser		
	110	115 120
Ala Pro Glu Lys Ala Arg Ala Phe Leu Pro Pro Ala Ser Ala Pro		
	125	130 135
Tyr Ser Leu Leu Asp Pro Ala Gln Phe Ser Ser Ser Ala Ile Gln		
	140	145 150
Lys Met Gly Glu Val Pro His Thr Leu Ala Thr Gly Ala Leu Pro		
	155	160 165
Tyr Ala Ser Thr Leu Gly Tyr Gln Asn Gly Ala Phe Gly Ser Leu		
	170	175 180
Ser Cys Pro Ser Gln His Thr His Thr His Pro Ser Pro Thr Asn		
	185	190 195
Pro Gly Tyr Val Val Pro Cys Asn Cys Thr Ala Trp Ser Ala Ser		
	200	205 210
Thr Leu Gln Pro Pro Val Ala Tyr Ile Leu Phe Pro Gly Met Thr		
	215	220 225
Lys Thr Gly Ile Asp Pro Tyr Ser Ser Ala His Ala Thr Ala Met		
	230	235 240

&lt;210&gt; 18

&lt;211&gt; 555



Glu Lys Pro Tyr	Glu Cys Lys Gln Cys Gly Lys Ala Phe Thr Trp	
	365	370 375
Ser Ser Thr Phe	Arg Glu His Val Arg Ile His Thr Gln Glu Gln	
	380	385 390
Leu Tyr Lys Cys	Glu Gln Cys Gly Lys Ala Phe Thr Ser Ser Arg	
	395	400 405
Ser Phe Arg Gly	His Leu Arg Thr His Thr Gly Glu Lys Pro Tyr	
	410	415 420
Glu Cys Lys Gln	Cys Gly Lys Thr Phe Thr Trp Ser Ser Thr Phe	
	425	430 435
Arg Glu His Val	Arg Ile His Thr Gln Glu Gln Leu His Lys Cys	
	440	445 450
Glu His Cys Gly	Lys Ala Phe Thr Ser Ser Arg Ala Phe Gln Gly	
	455	460 465
His Leu Arg Met	His Thr Gly Glu Lys Pro Tyr Glu Cys Lys Gln	
	470	475 480
Cys Gly Lys Thr	Phe Thr Trp Ser Ser Thr Leu His Asn His Val	
	485	490 495
Arg Met His Thr	Gly Glu Lys Pro His Lys Cys Lys Gln Cys Gly	
	500	505 510
Met Ser Phe Lys	Trp His Ser Ser Phe Arg Asn His Leu Arg Met	
	515	520 525
His Thr Gly Gln	Lys Ser His Glu Cys Gln Ser Tyr Ser Lys Ala	
	530	535 540
Phe Ser Cys Gln	Val Ile Leu Ser Lys Thr Ser Glu Ser Thr His	
	545	550 555

&lt;210&gt; 19

&lt;211&gt; 184

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7488247CD1

&lt;400&gt; 19

Met Ala Asp His Leu Met Leu Ala Glu Gly Tyr Arg Leu Val Gln	
1 5 10 15	
Arg Pro Pro Ser Ala Ala Ala Ala His Gly Pro His Ala Leu Arg	
20 25 30	
Thr Leu Pro Pro Tyr Ala Gly Pro Gly Leu Asp Ser Gly Leu Arg	
35 40 45	
Pro Arg Gly Ala Pro Leu Gly Pro Pro Pro Arg Gln Pro Gly	
50 55 60	
Ala Leu Ala Tyr Gly Ala Phe Gly Pro Pro Ser Ser Phe Gln Pro	
65 70 75	
Phe Pro Ala Val Pro Pro Ala Ala Gly Ile Ala His Leu Gln	
80 85 90	
Pro Val Ala Thr Pro Tyr Pro Gly Arg Ala Ala Ala Pro Pro Asn	
95 100 105	
Ala Pro Gly Gly Pro Pro Gly Pro Gln Pro Ala Pro Ser Ala Ala	
110 115 120	
Ala Pro Pro Pro Pro Ala His Ala Leu Gly Gly Met Asp Ala Glu	
125 130 135	

Leu Ile Asp Glu Glu Ala Leu Thr Ser Leu Glu Leu Glu Leu Gly  
                   140                  145                  150  
 Leu His Arg Val Arg Glu Leu Pro Glu Leu Phe Leu Gly Gln Ser  
                   155                  160                  165  
 Glu Phe Asp Cys Phe Ser Asp Leu Gly Ser Ala Pro Pro Ala Gly  
                   170                  175                  180  
 Ser Val Ser Cys

<210> 20

<211> 553

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1427269CD1

<400> 20

Met Pro Gly Met Met Glu Lys Gly Pro Glu Leu Leu Gly Lys Asn  
   1                  5                  10                  15  
 Arg Ser Ala Asn Gly Ser Ala Lys Ser Pro Ala Gly Gly Gly Gly  
                   20                  25                  30  
 Ser Gly Ala Ser Ser Thr Asn Gly Gly Leu His Tyr Ser Glu Pro  
                   35                  40                  45  
 Glu Ser Gly Cys Ser Ser Asp Asp Glu His Asp Val Gly Met Arg  
                   50                  55                  60  
 Val Gly Ala Glu Tyr Gln Ala Arg Ile Pro Glu Phe Asp Pro Gly  
                   65                  70                  75  
 Ala Thr Lys Tyr Thr Asp Lys Asp Asn Gly Gly Met Leu Val Trp  
                   80                  85                  90  
 Ser Pro Tyr His Ser Ile Pro Asp Ala Lys Leu Asp Glu Tyr Ile  
                   95                  100                 105  
 Ala Ile Ala Lys Glu Lys His Gly Tyr Asn Val Glu Gln Ala Leu  
                  110                 115                 120  
 Gly Met Leu Phe Trp His Lys His Asn Ile Glu Lys Ser Leu Ala  
                  125                 130                 135  
 Asp Leu Pro Asn Phe Thr Pro Phe Pro Asp Glu Trp Thr Val Glu  
                  140                 145                 150  
 Asp Lys Val Leu Phe Glu Gln Ala Phe Ser Phe His Gly Lys Ser  
                  155                 160                 165  
 Phe His Arg Ile Gln Gln Met Leu Pro Asp Lys Thr Ile Ala Ser  
                  170                 175                 180  
 Leu Val Lys Tyr Tyr Tyr Ser Trp Lys Lys Thr Arg Ser Arg Thr  
                  185                 190                 195  
 Ser Leu Met Asp Arg Gln Ala Arg Lys Leu Ala Asn Arg His Asn  
                  200                 205                 210  
 Gln Gly Asp Ser Asp Asp Asp Val Glu Glu Thr His Pro Met Asp  
                  215                 220                 225  
 Gly Asn Asp Ser Asp Tyr Asp Pro Lys Lys Glu Ala Lys Lys Glu  
                  230                 235                 240  
 Gly Asn Thr Glu Gln Pro Val Gln Thr Ser Lys Ile Gly Leu Gly  
                  245                 250                 255  
 Arg Arg Glu Tyr Gln Ser Leu Gln His Arg His His Ser Gln Arg  
                  260                 265                 270  
 Ser Lys Cys Arg Pro Pro Lys Gly Met Tyr Leu Thr Gln Glu Asp

	275		280		285
Val Val Ala Val	Ser Cys Ser Pro Asn	Ala Ala Asn Thr Ile	Leu		
	290		295		300
Arg Gln Leu Asp	Met Glu Leu Ile Ser	Leu Lys Arg Gln Val	Gln		
	305		310		315
Asn Ala Lys Gln	Val Asn Ser Ala Leu	Lys Gln Lys Met Glu	Gly		
	320		325		330
Gly Ile Glu Glu	Phe Lys Pro Pro Glu	Ser Asn Gln Lys Ile	Asn		
	335		340		345
Ala Arg Trp Thr	Thr Glu Glu Gln Leu	Leu Ala Val Gln Gly	Val		
	350		355		360
Arg Lys Tyr Gly	Lys Asp Phe Gln Ala	Ile Ala Asp Val Ile	Gly		
	365		370		375
Asn Lys Thr Val	Gly Gln Val Lys Asn	Phe Val Asn Tyr	Arg		
	380		385		390
Arg Arg Phe Asn	Leu Glu Glu Val Leu	Gln Glu Trp Glu Ala	Glu		
	395		400		405
Gln Gly Thr Gln	Ala Ser Asn Gly Asp	Ala Ser Thr Leu Gly	Glu		
	410		415		420
Glu Thr Lys Ser	Ala Ser Asn Val Pro	Ser Gly Lys Ser Thr	Asp		
	425		430		435
Glu Glu Glu Glu	Ala Gln Thr Pro Gln	Ala Pro Arg Thr Leu	Gly		
	440		445		450
Pro Ser Pro Pro	Ala Pro Ser Ser Thr	Pro Thr Pro Thr Ala	Pro		
	455		460		465
Ile Ala Thr Leu	Asn Gln Pro Pro Pro	Leu Leu Arg Pro Thr	Leu		
	470		475		480
Pro Ala Ala Pro	Ala Leu His Arg Gln	Pro Pro Pro Leu Gln	Gln		
	485		490		495
Gln Ala Arg Phe	Ile Gln Pro Arg Pro	Thr Leu Asn Gln Pro	Pro		
	500		505		510
Pro Pro Leu Ile	Arg Pro Ala Asn Ser	Met Pro Pro Arg Leu	Asn		
	515		520		525
Pro Arg Pro Val	Leu Ser Thr Val Gly	Gly Gln Gln Pro Pro	Ser		
	530		535		540
Leu Ile Gly Ile	Gln Thr Asp Ser Gln	Ser Ser Leu His			
	545		550		

&lt;210&gt; 21

&lt;211&gt; 371

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 103135CD1

&lt;400&gt; 21

Met Asp Met Ala	Gln Glu Pro Val Thr	Phe Arg Asp Val Ala	Ile
1	5	10	15
Tyr Phe Ser Arg	Glu Glu Trp Ala Cys	Leu Glu Pro Ser Gln	Arg
	20	25	30
Ala Leu Tyr Arg	Asp Val Met Leu Asp	Asn Phe Ser Ser Val	Ala
	35	40	45
Ala Leu Gly Phe	Cys Ser Pro Arg Pro	Asp Leu Val Ser Arg	Leu
	50	55	60

Glu Gln Trp Glu Glu Pro Trp Val Glu Asp Arg Glu Arg Pro Glu  
 65 70 75  
 Phe Gln Ala Val Gln Arg Gly Pro Arg Pro Gly Ala Arg Lys Ser  
 80 85 90  
 Ala Asp Pro Lys Arg His Cys Asp His Pro Ala Trp Ala His Lys  
 95 100 105  
 Lys Thr His Val Arg Arg Glu Arg Ala Arg Glu Gly Ser Ser Phe  
 110 115 120  
 Arg Lys Gly Phe Arg Leu Asp Thr Asp Asp Gly Gln Leu Pro Arg  
 125 130 135  
 Ala Ala Pro Glu Arg Thr Asp Ala Lys Pro Thr Ala Phe Pro Cys  
 140 145 150  
 Gln Val Leu Thr Gln Arg Cys Gly Arg Arg Pro Gly Arg Arg Glu  
 155 160 165  
 Arg Arg Lys Gln Arg Ala Val Glu Leu Ser Phe Ile Cys Gly Thr  
 170 175 180  
 Cys Gly Lys Ala Leu Ser Cys His Ser Arg Leu Leu Ala His Gln  
 185 190 195  
 Thr Val His Thr Gly Thr Lys Ala Phe Glu Cys Pro Glu Cys Gly  
 200 205 210  
 Gln Thr Phe Arg Trp Ala Ser Asn Leu Gln Arg His Gln Lys Asn  
 215 220 225  
 His Thr Arg Glu Lys Pro Phe Cys Cys Glu Ala Cys Gly Gln Ala  
 230 235 240  
 Phe Ser Leu Lys Asp Arg Leu Ala Gln His Arg Lys Val His Thr  
 245 250 255  
 Glu His Arg Pro Tyr Ser Cys Gly Asp Cys Gly Lys Ala Phe Lys  
 260 265 270  
 Gln Lys Ser Asn Leu Leu Arg His Gln Leu Val His Thr Gly Glu  
 275 280 285  
 Arg Pro Phe Tyr Cys Ala Asp Cys Gly Lys Ala Phe Arg Thr Lys  
 290 295 300  
 Glu Asn Leu Ser His His Gln Arg Val His Ser Gly Glu Lys Pro  
 305 310 315  
 Tyr Thr Cys Ala Glu Cys Gly Lys Ser Phe Arg Trp Pro Lys Gly  
 320 325 330  
 Phe Ser Ile His Arg Arg Leu His Leu Thr Lys Arg Phe Tyr Glu  
 335 340 345  
 Cys Gly His Cys Gly Lys Gly Phe Arg His Leu Gly Phe Phe Thr  
 350 355 360  
 Arg His Gln Arg Thr His Arg His Gly Glu Val  
 365 370

&lt;210&gt; 22

&lt;211&gt; 837

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1907346CD1

&lt;400&gt; 22

Met Leu Pro Lys Glu Glu Val Trp Lys Lys Arg Lys Arg Lys Glu  
 1 5 10 15  
 Lys Glu Ser Gly Met Ala Leu Thr Gln Val Arg Leu Thr Phe Arg

	20		25		30
Asp Val Ala Ile Glu Phe Ser Gln Glu Glu Trp Lys Cys Leu Asp					
	35		40		45
Pro Ala Gln Arg Ile Leu Tyr Arg Asp Val Met Leu Glu Asn Tyr					
	50		55		60
Trp Asn Leu Val Ser Leu Gly Leu Cys His Phe Asp Met Asn Ile					
	65		70		75
Ile Ser Met Leu Glu Glu Gly Lys Glu Pro Trp Thr Val Lys Ser					
	80		85		90
Cys Val Lys Ile Ala Arg Lys Pro Arg Thr Arg Glu Cys Val Lys					
	95		100		105
Gly Val Val Thr Asp Ile Pro Pro Lys Cys Thr Ile Lys Asp Leu					
	110		115		120
Leu Pro Lys Glu Lys Ser Ser Thr Glu Ala Val Phe His Thr Val					
	125		130		135
Val Leu Glu Arg His Glu Ser Pro Asp Ile Glu Asp Phe Ser Phe					
	140		145		150
Lys Glu Pro Gln Lys Asn Val His Asp Phe Glu Cys Gln Trp Arg					
	155		160		165
Asp Asp Thr Gly Asn Tyr Lys Gly Val Leu Met Ala Gln Lys Glu					
	170		175		180
Gly Lys Arg Asp Gln Arg Asp Arg Arg Asp Ile Glu Asn Lys Leu					
	185		190		195
Met Asn Asn Gln Leu Gly Val Ser Phe His Ser His Leu Pro Glu					
	200		205		210
Leu Gln Leu Phe Gln Gly Glu Gly Lys Met Tyr Glu Cys Asn Gln					
	215		220		225
Val Glu Lys Ser Thr Asn Asn Gly Ser Ser Val Ser Pro Leu Gln					
	230		235		240
Gln Ile Pro Ser Ser Val Gln Thr His Arg Ser Lys Lys Tyr His					
	245		250		255
Glu Leu Asn His Phe Ser Leu Leu Thr Gln Arg Arg Lys Ala Asn					
	260		265		270
Ser Cys Gly Lys Pro Tyr Lys Cys Asn Glu Cys Gly Lys Ala Phe					
	275		280		285
Thr Gln Asn Ser Asn Leu Thr Ser His Arg Arg Ile His Ser Gly					
	290		295		300
Glu Lys Pro Tyr Lys Cys Ser Glu Cys Gly Lys Thr Phe Thr Val					
	305		310		315
Arg Ser Asn Leu Thr Ile His Gln Val Ile His Thr Gly Glu Lys					
	320		325		330
Pro Tyr Lys Cys His Glu Cys Gly Lys Val Phe Arg His Asn Ser					
	335		340		345
Tyr Leu Ala Thr His Arg Arg Ile His Thr Gly Glu Lys Pro Tyr					
	350		355		360
Lys Cys Asn Glu Cys Gly Lys Ala Phe Arg Gly His Ser Asn Leu					
	365		370		375
Thr Thr His Gln Leu Ile His Thr Gly Glu Lys Pro Phe Lys Cys					
	380		385		390
Asn Glu Cys Gly Lys Leu Phe Thr Gln Asn Ser His Leu Ile Ser					
	395		400		405
His Trp Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Asn Glu					
	410		415		420
Cys Gly Lys Ala Phe Ser Val Arg Ser Ser Leu Ala Ile His Gln					
	425		430		435
Thr Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Asn Glu Cys Gly					

	440		445		450
Lys Val Phe Arg	Tyr Asn Ser Tyr Leu	Gly Arg His Arg Arg	Val		
	455		460		465
His Thr Gly Glu	Lys Pro Tyr Lys Cys	Asn Glu Cys Gly Lys	Ala		
	470		475		480
Phe Ser Met His	Ser Asn Leu Ala Thr	His Gln Val Ile His	Thr		
	485		490		495
Gly Thr Lys Pro	Phe Lys Cys Asn Glu	Cys Ser Lys Val Phe	Thr		
	500		505		510
Gln Asn Ser Gln	Leu Ala Asn His Arg	Arg Met His Thr Gly	Glu		
	515		520		525
Lys Thr Tyr Lys	Cys Asn Glu Cys Gly	Lys Ala Phe Ser Val	Arg		
	530		535		540
Ser Ser Leu Thr	Thr His Gln Ala Ile	His Ser Gly Glu Lys	Pro		
	545		550		555
Tyr Lys Cys Ile	Glu Cys Gly Lys Ser	Phe Thr Gln Lys Ser	His		
	560		565		570
Leu Arg Ser His	Arg Gly Ile His Ser	Gly Glu Lys Pro Tyr	Lys		
	575		580		585
Cys Asn Glu Cys	Gly Lys Val Phe Ala	Gln Thr Ser Gln Leu	Ala		
	590		595		600
Arg His Trp Arg	Val His Thr Gly Glu	Lys Pro Tyr Lys Cys	Asn		
	605		610		615
Asp Cys Gly Arg	Ala Phe Ser Asp Arg	Ser Ser Leu Thr Phe	His		
	620		625		630
Gln Ala Ile His	Thr Gly Glu Lys Pro	Tyr Lys Cys His Glu	Cys		
	635		640		645
Gly Lys Val Phe	Arg His Asn Ser Tyr	Leu Ala Thr His Arg	Arg		
	650		655		660
Ile His Thr Gly	Glu Lys Pro Tyr Lys	Cys Asn Glu Cys Gly	Lys		
	665		670		675
Ala Phe Ser Met	His Ser Asn Leu Thr	Thr His Lys Val Ile	His		
	680		685		690
Thr Gly Glu Lys	Pro Tyr Lys Cys Asn	Gln Cys Gly Lys Val	Phe		
	695		700		705
Thr Gln Asn Ser	His Leu Ala Asn His	Gln Arg Thr His Thr	Gly		
	710		715		720
Glu Lys Pro Tyr	Arg Cys Asn Glu Cys	Gly Lys Ala Phe Ser	Val		
	725		730		735
Arg Ser Ser Leu	Thr Thr His Gln Ala	Ile His Thr Gly Lys	Lys		
	740		745		750
Pro Tyr Lys Cys	Asn Glu Cys Gly Lys	Val Phe Thr Gln Asn	Ala		
	755		760		765
His Leu Ala Asn	His Arg Arg Ile His	Thr Gly Glu Lys Pro	Tyr		
	770		775		780
Arg Cys Thr Glu	Cys Gly Lys Ala Phe	Arg Val Arg Ser Ser	Leu		
	785		790		795
Thr Thr His Met	Ala Ile His Thr Gly	Glu Lys Arg Tyr Lys	Cys		
	800		805		810
Asn Glu Cys Gly	Lys Val Phe Arg Gln	Ser Ser Asn Leu Ala	Ser		
	815		820		825
His His Arg Met	His Thr Gly Glu Lys	Pro Tyr Lys			
	830		835		

&lt;210&gt; 23

&lt;211&gt; 549

<213> Homo sapiens

<221> misc feature

<223> Incyte ID No: 3041036CD1

Met	Ala	Ala	Gln	Leu	Leu	Thr	Asp	Glu	Ala	Leu	Glu	Ser	Val	Thr
1				5					10					15
Phe	Arg	Asp	Val	Thr	Val	Asp	Phe	Thr	Gln	Glu	Glu	Trp	Gln	Gln
				20					25					30
Leu	Glu	Pro	Ala	Gln	Lys	Asp	Leu	Tyr	Arg	Asp	Val	Met	Leu	Glu
				35					40					45
Asn	Tyr	Arg	Asn	Leu	Val	Ser	Leu	Asp	Trp	Glu	Thr	Arg	Pro	Glu
				50					55					60
Met	Lys	Glu	Leu	Asp	Pro	Lys	Asn	Asp	Ile	Ser	Glu	Asp	Lys	Leu
				65					70					75
Ser	Val	Val	Gly	Glu	Ala	Thr	Gly	Gly	Pro	Thr	Arg	Asn	Gly	Ala
				80					85					90
Arg	Gly	Pro	Gly	Ser	Glu	Gly	Val	Trp	Glu	Pro	Gly	Ser	Trp	Pro
				95					100					105
Glu	Arg	Pro	Arg	Gly	Asp	Ala	Gly	Ala	Glu	Trp	Glu	Pro	Leu	Gly
				110					115					120
Ile	Pro	Gln	Gly	Asn	Lys	Leu	Leu	Gly	Gly	Ser	Val	Pro	Ala	Cys
				125					130					135
His	Glu	Leu	Lys	Ala	Phe	Ala	Asn	Gln	Gly	Cys	Val	Leu	Val	Pro
				140					145					150
Pro	Arg	Leu	Asp	Asp	Pro	Thr	Glu	Lys	Gly	Ala	Cys	Pro	Pro	Val
				155					160					165
Arg	Arg	Gly	Lys	Asn	Phe	Ser	Ser	Thr	Ser	Asp	Leu	Ser	Lys	Pro
				170					175					180
Pro	Met	Pro	Cys	Glu	Glu	Lys	Lys	Thr	Tyr	Asp	Cys	Ser	Glu	Cys
				185					190					195
Gly	Lys	Ala	Phe	Ser	Arg	Ser	Ser	Ser	Leu	Ile	Lys	His	Gln	Arg
				200					205					210
Ile	His	Thr	Gly	Glu	Lys	Pro	Phe	Glu	Cys	Asp	Thr	Cys	Gly	Lys
				215					220					225
His	Phe	Ile	Glu	Arg	Ser	Ser	Leu	Thr	Ile	His	Gln	Arg	Val	His
				230					235					240
Thr	Gly	Glu	Lys	Pro	Tyr	Ala	Cys	Gly	Asp	Cys	Gly	Lys	Ala	Phe
				245					250					255
Ser	Gln	Arg	Met	Asn	Leu	Thr	Val	His	Gln	Arg	Thr	His	Thr	Gly
				260					265					270
Glu	Lys	Pro	Tyr	Val	Cys	Asp	Val	Cys	Gly	Lys	Ala	Phe	Arg	Lys
				275					280					285
Thr	Ser	Ser	Leu	Thr	Gln	His	Glu	Arg	Ile	His	Thr	Gly	Glu	Lys
				290					295					300
Pro	Tyr	Ala	Cys	Gly	Asp	Cys	Gly	Lys	Ala	Phe	Ser	Gln	Asn	Met
				305					310					315
His	Leu	Ile	Val	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr
				320					325					330
Val	Cys	Pro	Glu	Cys	Gly	Arg	Ala	Phe	Ser	Gln	Asn	Met	His	Leu
				335					340					345
Thr	Glu	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Ala	Cys
				350		</								

```

Lys Glu Cys Gly Lys Ala Phe Asn Lys Ser Ser Ser Leu Thr Leu
      365                      370                      375
His Gln Arg Asn His Thr Gly Glu Lys Pro Tyr Val Cys Gly Glu
      380                      385                      390
Cys Gly Lys Ala Phe Ser Gln Ser Ser Tyr Leu Ile Gln His Gln
      395                      400                      405
Arg Phe His Ile Gly Val Lys Pro Phe Glu Cys Ser Glu Cys Gly
      410                      415                      420
Lys Ala Phe Ser Lys Asn Ser Ser Leu Thr Gln His Gln Arg Ile
      425                      430                      435
His Thr Gly Glu Lys Pro Tyr Glu Cys Tyr Ile Cys Lys Lys His
      440                      445                      450
Phe Thr Gly Arg Ser Ser Leu Ile Val His Gln Ile Val His Thr
      455                      460                      465
Gly Glu Lys Pro Tyr Val Cys Gly Glu Cys Gly Lys Ala Phe Ser
      470                      475                      480
Gln Ser Ala Tyr Leu Ile Glu His Gln Arg Ile His Thr Gly Glu
      485                      490                      495
Lys Pro Tyr Arg Cys Gly Gln Cys Gly Lys Ser Phe Ile Lys Asn
      500                      505                      510
Ser Ser Leu Thr Val His Gln Arg Ile His Thr Gly Glu Lys Pro
      515                      520                      525
Tyr Arg Cys Gly Glu Cys Gly Lys Thr Phe Ser Arg Asn Thr Asn
      530                      535                      540
Leu Thr Arg His Leu Arg Ile His Thr
      545

```

&lt;210&gt; 24

&lt;211&gt; 555

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3856879CD1

&lt;400&gt; 24

```

Met Ala Ala Ala Arg Leu Leu Pro Val Pro Ala Gly Pro Gln Ala
  1          5          10          15
Lys Leu Thr Phe Glu Asp Val Ala Val Leu Leu Ser Gln Asp Glu
      20          25          30
Trp Asp Arg Leu Cys Pro Ala Gln Arg Gly Leu Tyr Arg Asn Val
      35          40          45
Met Met Glu Thr Tyr Gly Asn Val Val Ser Leu Gly Leu Pro Gly
      50          55          60
Ser Lys Pro Asp Ile Ile Ser Gln Leu Glu Arg Gly Glu Asp Pro
      65          70          75
Trp Val Leu Asp Arg Lys Gly Ala Lys Lys Ser Gln Gly Leu Trp
      80          85          90
Ser Asp Tyr Ser Asp Asn Leu Lys Tyr Asp His Thr Thr Ala Cys
      95          100         105
Thr Gln Gln Asp Ser Leu Ser Cys Pro Trp Glu Cys Glu Thr Lys
      110         115         120
Gly Glu Ser Gln Asn Thr Asp Leu Ser Pro Lys Pro Leu Ile Ser
      125         130         135
Glu Gln Thr Val Ile Leu Gly Lys Thr Pro Leu Gly Arg Ile Asp

```

	140		145		150
Gln Glu Asn Asn	Glu Thr Lys Gln Ser	Phe Cys Leu Ser Pro	Asn		
	155		160		165
Ser Val Asp His	Arg Glu Val Gln Val	Leu Ser Gln Ser Met	Pro		
	170		175		180
Leu Thr Pro His	Gln Ala Val Pro Ser	Gly Glu Arg Pro Tyr	Met		
	185		190		195
Cys Val Glu Cys	Gly Lys Cys Phe Gly	Arg Ser Ser His Leu	Leu		
	200		205		210
Gln His Gln Arg	Ile His Thr Gly Glu	Lys Pro Tyr Val Cys	Ser		
	215		220		225
Val Cys Gly Lys	Ala Phe Ser Gln Ser	Ser Val Leu Ser Lys	His		
	230		235		240
Arg Arg Ile His	Thr Gly Glu Lys Pro	Tyr Glu Cys Asn Glu	Cys		
	245		250		255
Gly Lys Ala Phe	Arg Val Ser Ser Asp	Leu Ala Gln His His	Lys		
	260		265		270
Ile His Thr Gly	Glu Lys Pro His Glu	Cys Leu Glu Cys Arg	Lys		
	275		280		285
Ala Phe Thr Gln	Leu Ser His Leu Ile	Gln His Gln Arg Ile	His		
	290		295		300
Thr Gly Glu Arg	Pro Tyr Val Cys Pro	Leu Cys Gly Lys Ala	Phe		
	305		310		315
Asn His Ser Thr	Val Leu Arg Ser His	Gln Arg Val His Thr	Gly		
	320		325		330
Glu Lys Pro His	Arg Cys Asn Glu Cys	Gly Lys Thr Phe Ser	Val		
	335		340		345
Lys Arg Thr Leu	Leu Gln His Gln Arg	Ile His Thr Gly Glu	Lys		
	350		355		360
Pro Tyr Thr Cys	Ser Glu Cys Gly Lys	Ala Phe Ser Asp Arg	Ser		
	365		370		375
Val Leu Ile Gln	His His Asn Val His	Thr Gly Glu Lys Pro	Tyr		
	380		385		390
Glu Cys Ser Glu	Cys Gly Lys Thr Phe	Ser His Arg Ser Thr	Leu		
	395		400		405
Met Asn His Glu	Arg Ile His Thr Glu	Glu Lys Pro Tyr Ala	Cys		
	410		415		420
Tyr Glu Cys Gly	Lys Ala Phe Val Gln	His Ser His Leu Ile	Gln		
	425		430		435
His Gln Arg Val	His Thr Gly Glu Lys	Pro Tyr Val Cys Gly	Glu		
	440		445		450
Cys Gly His Ala	Phe Ser Ala Arg Arg	Ser Leu Ile Gln His	Glu		
	455		460		465
Arg Ile His Thr	Gly Glu Lys Pro Phe	Gln Cys Thr Glu Cys	Gly		
	470		475		480
Lys Ala Phe Ser	Leu Lys Ala Thr Leu	Ile Val His Leu Arg	Thr		
	485		490		495
His Thr Gly Glu	Lys Pro Tyr Glu Cys	Asn Ser Cys Gly Lys	Ala		
	500		505		510
Phe Ser Gln Tyr	Ser Val Leu Ile Gln	His Gln Arg Ile His	Thr		
	515		520		525
Gly Glu Lys Pro	Tyr Glu Cys Gly Glu	Cys Gly Arg Ala Phe	Asn		
	530		535		540
Gln His Gly His	Leu Ile Gln His Gln	Lys Val His Arg Lys	Leu		
	545		550		555

<210> 25  
 <211> 601  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 4178665CD1

<400> 25  
 Met Leu Cys Trp Leu Gln Glu Asn Asn Phe Cys Leu Leu Leu Cys  
 1 5 10 15  
 Phe Leu Ser Gly Leu Leu Ser Arg His Lys Thr Lys Lys Leu Ser  
 20 25 30  
 Ser Glu Lys Asp Ile His Glu Ile Ser Leu Ser Lys Glu Ser Ile  
 35 40 45  
 Ile Glu Lys Ser Lys Thr Leu Arg Leu Lys Gly Ser Ile Phe Arg  
 50 55 60  
 Asn Glu Trp Gln Asn Lys Ser Glu Phe Glu Gly Gln Gln Gly Leu  
 65 70 75  
 Lys Glu Arg Ser Ile Ser Gln Lys Lys Ile Val Ser Lys Lys Met  
 80 85 90  
 Ser Thr Asp Arg Lys Arg Pro Ser Phe Thr Leu Asn Gln Arg Ile  
 95 100 105  
 His Asn Ser Glu Lys Ser Cys Asp Ser His Leu Val Gln His Gly  
 110 115 120  
 Lys Ile Asp Ser Asp Val Lys His Asp Cys Lys Glu Cys Gly Ser  
 125 130 135  
 Thr Phe Asn Asn Val Tyr Gln Leu Thr Leu His Gln Lys Ile His  
 140 145 150  
 Thr Gly Glu Lys Ser Cys Lys Cys Glu Lys Cys Gly Lys Val Phe  
 155 160 165  
 Ser His Ser Tyr Gln Leu Thr Leu His Gln Arg Phe His Thr Gly  
 170 175 180  
 Glu Lys Pro Tyr Glu Cys Gln Glu Cys Gly Lys Thr Phe Thr Leu  
 185 190 195  
 Tyr Pro Gln Leu Asn Arg His Gln Lys Ile His Thr Gly Lys Lys  
 200 205 210  
 Pro Tyr Met Cys Lys Lys Cys Asp Lys Gly Phe Phe Ser Arg Leu  
 215 220 225  
 Glu Leu Thr Gln His Lys Arg Ile His Thr Gly Lys Lys Ser Tyr  
 230 235 240  
 Glu Cys Lys Glu Cys Gly Lys Val Phe Gln Leu Ile Phe Tyr Phe  
 245 250 255  
 Lys Glu His Glu Arg Ile His Thr Gly Lys Lys Pro Tyr Glu Cys  
 260 265 270  
 Lys Glu Cys Gly Lys Ala Phe Ser Val Cys Gly Gln Leu Thr Arg  
 275 280 285  
 His Gln Lys Ile His Thr Gly Val Lys Pro Tyr Glu Cys Lys Glu  
 290 295 300  
 Cys Gly Lys Thr Phe Arg Leu Ser Phe Tyr Leu Thr Glu His Arg  
 305 310 315  
 Arg Thr His Ala Gly Lys Lys Pro Tyr Glu Cys Lys Glu Cys Gly  
 320 325 330  
 Lys Ser Phe Asn Val Arg Gly Gln Leu Asn Arg His Lys Thr Ile  
 335 340 345

```

His Thr Gly Ile Lys Pro Phe Ala Cys Lys Val Cys Glu Lys Ala
      350                      355                      360
Phe Ser Tyr Ser Gly Asp Leu Arg Val His Ser Arg Ile His Thr
      365                      370                      375
Gly Glu Lys Pro Tyr Glu Cys Lys Glu Cys Gly Lys Ala Phe Met
      380                      385                      390
Leu Arg Ser Val Leu Thr Glu His Gln Arg Leu His Thr Gly Val
      395                      400                      405
Lys Pro Tyr Glu Cys Lys Glu Cys Gly Lys Thr Phe Arg Val Arg
      410                      415                      420
Ser Gln Ile Ser Leu His Lys Lys Ile His Thr Asp Val Lys Pro
      425                      430                      435
Tyr Lys Cys Val Arg Cys Gly Lys Thr Phe Arg Phe Gly Phe Tyr
      440                      445                      450
Leu Thr Glu His Gln Arg Ile His Thr Gly Glu Lys Pro Tyr Lys
      455                      460                      465
Cys Lys Glu Cys Gly Lys Ala Phe Ile Arg Arg Gly Asn Leu Lys
      470                      475                      480
Glu His Leu Lys Ile His Ser Gly Leu Lys Pro Tyr Asp Cys Lys
      485                      490                      495
Glu Cys Gly Lys Ser Phe Ser Arg Arg Gly Gln Phe Thr Glu His
      500                      505                      510
Gln Lys Ile His Thr Gly Val Lys Pro Tyr Lys Cys Lys Glu Cys
      515                      520                      525
Gly Lys Ala Phe Ser Arg Ser Val Asp Leu Arg Ile His Gln Arg
      530                      535                      540
Ile His Thr Gly Glu Lys Pro Tyr Glu Cys Lys Gln Cys Gly Lys
      545                      550                      555
Ala Phe Arg Leu Asn Ser His Leu Thr Glu His Gln Arg Ile His
      560                      565                      570
Thr Gly Glu Lys Pro Tyr Glu Cys Lys Val Cys Arg Lys Ala Phe
      575                      580                      585
Arg Gln Tyr Ser His Leu Tyr Gln His Gln Lys Thr His Asn Val
      590                      595                      600
Ile

```

&lt;210&gt; 26

&lt;211&gt; 743

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7493326CD1

&lt;400&gt; 26

```

Met Met Gln Ala Gln Glu Ser Leu Thr Leu Glu Asp Val Ala Val
  1          5          10          15
Asp Phe Thr Trp Glu Glu Trp Gln Phe Leu Ser Pro Ala Gln Lys
      20          25          30
Asp Leu Tyr Arg Asp Val Met Leu Glu Asn Tyr Ser Asn Leu Val
      35          40          45
Ala Val Gly Tyr Gln Ala Ser Lys Pro Asp Ala Leu Ser Lys Leu
      50          55          60
Glu Arg Gly Glu Glu Thr Cys Thr Thr Glu Asp Glu Ile Tyr Ser

```

	65		70		75
Arg Ile Cys Ser Asp	Ser Gly Gly Ala	Ser Gly Gly Ala Tyr Ala			
	80		85		90
Glu Ile Arg Lys Ile	Asp Asp Pro Leu	Gln His His Leu Gln Asn			
	95		100		105
Gln Ser Ile Gln Lys	Ser Val Lys Gln	Cys His Glu Gln Asn Met			
	110		115		120
Phe Gly Asn Ile Val	Asn Gln Asn Lys	Gly His Phe Leu Leu Lys			
	125		130		135
Gln Asp Cys Asp Thr	Phe Asp Leu His	Glu Lys Pro Leu Lys Ser			
	140		145		150
Asn Leu Ser Phe Glu	Asn Gln Lys Arg	Ser Ser Gly Leu Lys Asn			
	155		160		165
Ser Ala Glu Phe Asn	Arg Asp Gly Lys	Ser Leu Phe His Ala Asn			
	170		175		180
His Lys Gln Phe Tyr	Thr Glu Met Lys	Phe Pro Ala Ile Ala Lys			
	185		190		195
Pro Ile Asn Lys Ser	Gln Phe Ile Lys	Gln Gln Arg Thr His Asn			
	200		205		210
Ile Glu Asn Ala His	Val Cys Ser Glu	Cys Gly Lys Ala Phe Leu			
	215		220		225
Lys Leu Ser Gln Phe	Ile Asp His Gln	Arg Val His Thr Gly Glu			
	230		235		240
Lys Pro His Val Cys	Ser Met Cys Gly	Lys Ala Phe Ser Arg Lys			
	245		250		255
Ser Arg Leu Met Asp	His Gln Arg Thr	His Thr Glu Leu Lys His			
	260		265		270
Tyr Glu Cys Thr Glu	Cys Asp Lys Thr	Phe Leu Lys Lys Ser Gln			
	275		280		285
Leu Asn Ile His Gln	Lys Thr His Met	Gly Gly Lys Pro Tyr Thr			
	290		295		300
Cys Ser Gln Cys Gly	Lys Ala Phe Ile	Lys Lys Cys Arg Leu Ile			
	305		310		315
Tyr His Gln Arg Thr	His Thr Gly Glu	Lys Pro His Gly Cys Ser			
	320		325		330
Val Cys Gly Lys Ala	Phe Ser Thr Lys	Phe Ser Leu Thr Thr His			
	335		340		345
Gln Lys Thr His Thr	Gly Glu Lys Pro	Tyr Ile Cys Ser Glu Cys			
	350		355		360
Gly Lys Gly Phe Ile	Glu Lys Arg Arg	Leu Thr Ala His His Arg			
	365		370		375
Thr His Thr Gly Glu	Lys Pro Phe Ile	Cys Asn Lys Cys Gly Lys			
	380		385		390
Gly Phe Thr Leu Lys	Asn Ser Leu Ile	Thr His Gln Gln Thr His			
	395		400		405
Thr Gly Glu Lys Leu	Tyr Thr Cys Ser	Glu Cys Gly Lys Gly Phe			
	410		415		420
Ser Met Lys His Cys	Leu Met Val His	Gln Arg Thr His Thr Gly			
	425		430		435
Glu Lys Pro Tyr Lys	Cys Asn Glu Cys	Gly Lys Gly Phe Ala Leu			
	440		445		450
Lys Ser Pro Leu Ile	Arg His Gln Arg	Thr His Thr Gly Glu Lys			
	455		460		465
Pro Tyr Val Cys Thr	Glu Cys Arg Lys	Gly Phe Thr Met Lys Ser			
	470		475		480
Asp Leu Ile Val His	Gln Arg Thr His	Thr Ala Glu Lys Pro Tyr			

	485		490		495
Ile Cys Asn Asp Cys Gly Lys Gly Phe Thr Val Lys Ser Arg Leu					
	500		505		510
Ile Val His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Val Cys					
	515		520		525
Gly Glu Cys Gly Lys Gly Phe Pro Ala Lys Ile Arg Leu Met Gly					
	530		535		540
His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Ile Cys Asn Glu					
	545		550		555
Cys Gly Lys Gly Phe Thr Glu Lys Ser His Leu Asn Val His Arg					
	560		565		570
Arg Thr His Thr Gly Glu Lys Pro Tyr Val Cys Ser Glu Cys Gly					
	575		580		585
Lys Gly Leu Leu Gly Arg Ala Cys Ser Leu His His Gln Ala Asn					
	590		595		600
Ser Tyr Trp Gly Glu Lys Pro Tyr Ile Cys Asn Glu Cys Gly Lys					
	605		610		615
Gly Phe Ser Met Lys Ser Thr Leu Ser Ile His Gln Gln Thr His					
	620		625		630
Thr Gly Glu Lys Pro Tyr Lys Cys Asn Glu Cys Asp Lys Thr Phe					
	635		640		645
Arg Lys Lys Thr Cys Leu Ile Gln His Gln Arg Phe His Thr Gly					
	650		655		660
Lys Thr Ser Phe Ala Cys Thr Glu Cys Gly Lys Phe Ser Leu Arg					
	665		670		675
Lys Asn Asp Leu Ile Thr His Gln Arg Ile His Thr Gly Glu Lys					
	680		685		690
Pro Tyr Lys Cys Ser Asp Cys Gly Lys Ala Phe Thr Thr Lys Ser					
	695		700		705
Gly Leu Asn Val His Gln Arg Lys His Thr Gly Glu Arg Pro Tyr					
	710		715		720
Gly Cys Ser Asp Cys Gly Lys Ala Phe Ala His Leu Ser Ile Leu					
	725		730		735
Val Lys His Lys Arg Ile His Arg					
	740				

&lt;210&gt; 27

&lt;211&gt; 490

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1553836CD1

&lt;400&gt; 27

Met Lys Met Arg Arg Ile Lys Pro Ala Ala Thr Ser His Val Glu		
1	5	10
Gly Ser Gly Gly Val Ser Ala Lys Gly Lys Arg Lys Pro Arg Gln		
	20	25
Glu Glu Asp Glu Asp Tyr Arg Glu Phe Pro Gln Lys Lys His Lys		
	35	40
Leu Tyr Gly Arg Lys Gln Arg Pro Lys Thr Gln Pro Asn Pro Lys		
	50	55
Ser Gln Ala Arg Arg Ile Arg Lys Glu Pro Pro Val Tyr Ala Ala		
	65	70
		75

Gly	Ser	Leu	Glu	Glu	Gln	Trp	Tyr	Leu	Glu	Ile	Val	Asp	Lys	Gly	80	85	90
Ser	Val	Ser	Cys	Pro	Thr	Cys	Gln	Ala	Val	Gly	Arg	Lys	Thr	Ile	95	100	105
Glu	Gly	Leu	Lys	Lys	His	Met	Glu	Asn	Cys	Lys	Gln	Glu	Met	Phe	110	115	120
Thr	Cys	His	His	Cys	Gly	Lys	Gln	Leu	Arg	Ser	Leu	Ala	Gly	Met	125	130	135
Lys	Tyr	His	Val	Met	Ala	Asn	His	Asn	Ser	Leu	Pro	Ile	Leu	Lys	140	145	150
Ala	Gly	Asp	Glu	Ile	Asp	Glu	Pro	Ser	Glu	Arg	Glu	Arg	Leu	Arg	155	160	165
Thr	Val	Leu	Lys	Arg	Leu	Gly	Lys	Leu	Arg	Cys	Met	Arg	Glu	Ser	170	175	180
Cys	Ser	Ser	Ser	Phe	Thr	Ser	Ile	Met	Gly	Tyr	Leu	Tyr	His	Val	185	190	195
Arg	Lys	Cys	Gly	Lys	Gly	Ala	Ala	Glu	Leu	Glu	Lys	Met	Thr	Leu	200	205	210
Lys	Cys	His	His	Cys	Gly	Lys	Pro	Tyr	Arg	Ser	Lys	Ala	Gly	Leu	215	220	225
Ala	Tyr	His	Leu	Arg	Ser	Glu	His	Gly	Pro	Ile	Ser	Phe	Phe	Pro	230	235	240
Glu	Ser	Gly	Gln	Pro	Glu	Cys	Leu	Lys	Glu	Met	Asn	Leu	Glu	Ser	245	250	255
Lys	Ser	Gly	Gly	Arg	Val	Gln	Arg	Arg	Ser	Ala	Lys	Ile	Ala	Val	260	265	270
Tyr	His	Leu	Gln	Glu	Leu	Ala	Ser	Ala	Glu	Leu	Ala	Lys	Glu	Trp	275	280	285
Pro	Lys	Arg	Lys	Val	Leu	Gln	Asp	Leu	Val	Pro	Asp	Asp	Arg	Lys	290	295	300
Leu	Lys	Tyr	Thr	Arg	Pro	Gly	Leu	Pro	Thr	Phe	Ser	Gln	Glu	Val	305	310	315
Leu	His	Lys	Trp	Lys	Thr	Asp	Ile	Lys	Lys	Tyr	His	Arg	Ile	Gln	320	325	330
Cys	Pro	Asn	Gln	Gly	Cys	Glu	Ala	Val	Tyr	Ser	Ser	Val	Ser	Gly	335	340	345
Leu	Lys	Ala	His	Leu	Gly	Ser	Cys	Thr	Leu	Gly	Asn	Phe	Val	Ala	350	355	360
Gly	Lys	Tyr	Lys	Cys	Leu	Leu	Cys	Gln	Lys	Glu	Phe	Val	Ser	Glu	365	370	375
Ser	Gly	Val	Lys	Tyr	His	Ile	Asn	Ser	Val	His	Ala	Glu	Asp	Trp	380	385	390
Phe	Val	Val	Asn	Pro	Thr	Thr	Thr	Lys	Ser	Phe	Glu	Lys	Leu	Met	395	400	405
Lys	Ile	Lys	Gln	Arg	Gln	Gln	Glu	Glu	Glu	Lys	Arg	Arg	Gln	Gln	410	415	420
His	Arg	Ser	Arg	Arg	Ser	Leu	Arg	Arg	Arg	Gln	Gln	Pro	Gly	Ile	425	430	435
Glu	Leu	Pro	Glu	Thr	Glu	Leu	Ser	Leu	Arg	Val	Gly	Lys	Asp	Gln	440	445	450
Arg	Arg	Asn	Asn	Glu	Glu	Leu	Val	Val	Ser	Ala	Ser	Cys	Lys	Glu	455	460	465
Pro	Glu	Gln	Glu	Pro	Val	Pro	Ala	Gln	Phe	Gln	Lys	Val	Lys	Pro	470	475	480
Pro	Lys	Thr	Asn	His	Lys	Arg	Gly	Arg	Lys						485	490	

<210> 28  
 <211> 665  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1908201CD1

<400> 28  
 Met Pro Leu Arg Asp Lys Tyr Cys Gln Thr Asp His His His His  
 1 5 10 15  
 Gly Cys Cys Glu Pro Val Tyr Ile Leu Glu Pro Gly Asp Pro Pro  
 20 25 30  
 Leu Leu Gln Gln Pro Leu Gln Thr Ser Lys Ser Gly Ile Gln Gln  
 35 40 45  
 Ile Ile Glu Cys Phe Arg Ser Gly Thr Lys Gln Leu Lys His Ile  
 50 55 60  
 Leu Leu Lys Asp Val Asp Thr Ile Phe Glu Cys Lys Leu Cys Arg  
 65 70 75  
 Ser Leu Phe Arg Gly Leu Pro Asn Leu Ile Thr His Lys Lys Phe  
 80 85 90  
 Tyr Cys Pro Pro Ser Leu Gln Met Asp Asp Asn Leu Pro Asp Val  
 95 100 105  
 Asn Asp Lys Gln Ser Gln Ala Ile Asn Asp Leu Leu Glu Ala Ile  
 110 115 120  
 Tyr Pro Ser Val Asp Lys Arg Glu Tyr Ile Ile Lys Leu Glu Pro  
 125 130 135  
 Ile Glu Thr Asn Gln Asn Ala Val Phe Gln Tyr Ile Ser Arg Thr  
 140 145 150  
 Asp Asn Pro Ile Glu Val Thr Glu Ser Ser Ser Thr Pro Glu Gln  
 155 160 165  
 Thr Glu Val Gln Ile Gln Glu Thr Ser Thr Glu Gln Ser Lys Thr  
 170 175 180  
 Val Pro Val Thr Asp Thr Glu Val Glu Thr Val Glu Pro Pro Pro  
 185 190 195  
 Val Glu Ile Val Thr Asp Glu Val Ala Pro Thr Ser Asp Glu Gln  
 200 205 210  
 Pro Gln Glu Ser Gln Ala Asp Leu Glu Thr Ser Asp Asn Ser Asp  
 215 220 225  
 Phe Gly His Gln Leu Ile Cys Cys Leu Cys Arg Lys Glu Phe Asn  
 230 235 240  
 Ser Arg Arg Gly Val Arg Arg His Ile Arg Lys Val His Lys Lys  
 245 250 255  
 Lys Met Glu Glu Leu Lys Lys Tyr Ile Glu Thr Arg Lys Asn Pro  
 260 265 270  
 Asn Gln Ser Ser Lys Gly Arg Ser Lys Asn Val Leu Val Pro Leu  
 275 280 285  
 Ser Arg Ser Cys Pro Val Cys Cys Lys Ser Phe Ala Thr Lys Ala  
 290 295 300  
 Asn Val Arg Arg His Phe Asp Glu Val His Arg Gly Leu Arg Arg  
 305 310 315  
 Asp Ser Ile Thr Pro Asp Ile Ala Thr Lys Pro Gly Gln Pro Leu  
 320 325 330  
 Phe Leu Asp Ser Ile Ser Pro Lys Lys Ser Phe Lys Thr Arg Lys  
 335 340 345

Gln Lys Ser Ser Ser Lys Ala Glu Tyr Asn Leu Thr Ala Cys Lys  
 350 355 360  
 Cys Leu Leu Cys Lys Arg Lys Tyr Ser Ser Gln Ile Met Leu Lys  
 365 370 375  
 Arg His Met Gln Ile Val His Lys Ile Thr Leu Ser Gly Thr Asn  
 380 385 390  
 Ser Lys Arg Glu Lys Gly Pro Asn Asn Thr Ala Asn Ser Ser Glu  
 395 400 405  
 Ile Lys Val Lys Val Glu Pro Ala Asp Ser Val Glu Ser Ser Pro  
 410 415 420  
 Pro Ser Ile Thr His Ser Pro Gln Asn Glu Leu Lys Gly Thr Asn  
 425 430 435  
 His Ser Asn Glu Lys Lys Asn Thr Pro Ala Ala Gln Lys Asn Lys  
 440 445 450  
 Val Lys Gln Asp Ser Glu Ser Pro Lys Ser Thr Ser Pro Ser Ala  
 455 460 465  
 Ala Gly Gly Gln Gln Lys Thr Arg Lys Pro Lys Leu Ser Ala Gly  
 470 475 480  
 Phe Asp Phe Lys Gln Leu Tyr Cys Lys Leu Cys Lys Arg Gln Phe  
 485 490 495  
 Thr Ser Lys Gln Asn Leu Thr Lys His Ile Glu Leu His Thr Asp  
 500 505 510  
 Gly Asn Asn Ile Tyr Val Lys Phe Tyr Lys Cys Pro Leu Cys Thr  
 515 520 525  
 Tyr Glu Thr Arg Arg Lys Arg Asp Val Ile Arg His Ile Thr Val  
 530 535 540  
 Val His Lys Lys Ser Ser Arg Tyr Leu Gly Lys Ile Thr Ala Ser  
 545 550 555  
 Leu Glu Ile Arg Ala Ile Lys Lys Pro Ile Asp Phe Val Leu Asn  
 560 565 570  
 Lys Val Ala Lys Arg Gly Pro Ser Arg Asp Glu Ala Lys His Ser  
 575 580 585  
 Asp Ser Lys His Asp Gly Thr Ser Asn Ser Pro Ser Lys Lys Tyr  
 590 595 600  
 Glu Val Ala Asp Val Gly Ile Glu Val Lys Val Thr Lys Asn Phe  
 605 610 615  
 Ser Leu His Arg Cys Asn Lys Cys Gly Lys Ala Phe Ala Lys Lys  
 620 625 630  
 Thr Tyr Leu Glu His His Lys Lys Thr His Lys Ala Asn Ala Ser  
 635 640 645  
 Asn Ser Pro Glu Gly Asn Lys Thr Lys Gly Arg Ser Thr Arg Ser  
 650 655 660  
 Lys Ala Leu Val Trp  
 665

&lt;210&gt; 29

&lt;211&gt; 570

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2827615CD1

&lt;400&gt; 29

Met Ser Lys Asp Leu Val Thr Phe Gly Asp Val Ala Val Asn Phe

1	5	10	15
Ser Gln Glu Glu Trp	Glu Trp Leu Asn Pro	Ala Gln Arg Asn Leu	
20	25	30	
Tyr Arg Lys Val Met	Leu Glu Asn Tyr Arg	Ser Leu Val Ser Leu	
35	40	45	
Ala Gly Val Ser Val	Ser Lys Pro Asp Val	Ile Ser Leu Leu Glu	
50	55	60	
Gln Gly Lys Glu Pro	Trp Met Val Lys Lys	Glu Gly Thr Arg Gly	
65	70	75	
Pro Cys Pro Asp Trp	Glu Tyr Val Phe Lys	Asn Ser Glu Phe Ser	
80	85	90	
Ser Lys Gln Glu Thr	Tyr Glu Glu Ser Ser	Lys Val Val Thr Val	
95	100	105	
Gly Ala Arg His Leu	Ser Tyr Ser Leu Asp	Tyr Pro Ser Leu Arg	
110	115	120	
Glu Asp Cys Gln Ser	Glu Asp Trp Tyr Lys	Asn Gln Leu Gly Ser	
125	130	135	
Gln Glu Val His Leu	Ser Gln Leu Ile Ile	Thr His Lys Glu Ile	
140	145	150	
Leu Pro Glu Val Gln	Asn Lys Glu Tyr Asn	Lys Ser Trp Gln Thr	
155	160	165	
Phe His Gln Asp Thr	Ile Phe Asp Ile Gln	Gln Ser Phe Pro Thr	
170	175	180	
Lys Glu Lys Ala His	Lys His Glu Pro Gln	Lys Lys Ser Tyr Arg	
185	190	195	
Lys Lys Ser Val Glu	Met Lys His Arg Lys	Val Tyr Val Glu Lys	
200	205	210	
Lys Leu Leu Lys Cys	Asn Asp Cys Glu Lys	Val Phe Asn Gln Ser	
215	220	225	
Ser Ser Leu Thr Leu	His Gln Arg Ile His	Thr Gly Glu Lys Pro	
230	235	240	
Tyr Ala Cys Val Glu	Cys Gly Lys Thr Phe	Ser Gln Ser Ala Asn	
245	250	255	
Leu Ala Gln His Lys	Arg Ile His Thr Gly	Glu Lys Pro Tyr Glu	
260	265	270	
Cys Lys Glu Cys Arg	Lys Ala Phe Ser Gln	Asn Ala His Leu Ala	
275	280	285	
Gln His Gln Arg Val	His Thr Gly Glu Lys	Pro Tyr Gln Cys Lys	
290	295	300	
Glu Cys Lys Lys Ala	Phe Ser Gln Ile Ala	His Leu Thr Gln His	
305	310	315	
Gln Arg Val His Thr	Gly Glu Arg Pro Phe	Glu Cys Ile Glu Cys	
320	325	330	
Gly Lys Ala Phe Ser	Asn Gly Ser Phe Leu	Ala Gln His Gln Arg	
335	340	345	
Ile His Thr Gly Glu	Lys Pro Tyr Val Cys	Asn Val Cys Gly Lys	
350	355	360	
Ala Phe Ser His Arg	Gly Tyr Leu Ile Val	His Gln Arg Ile His	
365	370	375	
Thr Gly Glu Arg Pro	Tyr Glu Cys Lys Glu	Cys Arg Lys Ala Phe	
380	385	390	
Ser Gln Tyr Ala His	Leu Ala Gln His Gln	Arg Val His Thr Gly	
395	400	405	
Glu Lys Pro Tyr Glu	Cys Lys Val Cys Arg	Lys Ala Phe Ser Gln	
410	415	420	
Ile Ala Tyr Leu Asp	Gln His Gln Arg Val	His Thr Gly Glu Lys	

	425		430		435
Pro Tyr Glu Cys Ile Glu Cys Gly Lys Ala Phe Ser Asn Ser Ser					
	440		445		450
Ser Leu Ala Gln His Gln Arg Ser His Thr Gly Glu Lys Pro Tyr					
	455		460		465
Met Cys Lys Glu Cys Arg Lys Thr Phe Ser Gln Asn Ala Gly Leu					
	470		475		480
Ala Gln His Gln Arg Ile His Thr Gly Glu Lys Pro Tyr Glu Cys					
	485		490		495
Asn Val Cys Gly Lys Ala Phe Ser Tyr Ser Gly Ser Leu Thr Leu					
	500		505		510
His Gln Arg Ile His Thr Gly Glu Arg Pro Tyr Glu Cys Lys Asp					
	515		520		525
Cys Arg Lys Ser Phe Arg Gln Arg Ala His Leu Ala His His Glu					
	530		535		540
Arg Ile His Thr Met Glu Ser Phe Leu Thr Leu Ser Ser Pro Ser					
	545		550		555
Pro Ser Thr Ser Asn Gln Leu Pro Arg Pro Val Gly Phe Ile Ser					
	560		565		570

&lt;210&gt; 30

&lt;211&gt; 1712

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4304550CD1

&lt;400&gt; 30

Met Glu Arg Asn Val Leu Thr Thr Phe Ser Gln Glu Met Ser Gln		
1	5	15
Leu Ile Leu Asn Glu Met Pro Lys Ala Glu Tyr Ser Ser Leu Phe		
	20	30
Asn Asp Phe Val Glu Ser Glu Phe Phe Leu Ile Asp Gly Asp Ser		
	35	45
Leu Leu Ile Thr Cys Ile Cys Glu Ile Ser Phe Lys Pro Gly Gln		
	50	60
Asn Leu His Phe Phe Tyr Leu Val Glu Arg Tyr Leu Val Asp Leu		
	65	75
Ile Ser Lys Gly Gly Gln Phe Thr Ile Val Phe Phe Lys Asp Ala		
	80	90
Glu Tyr Ala Tyr Phe Asn Phe Pro Glu Leu Leu Ser Leu Arg Thr		
	95	105
Ala Leu Ile Leu His Leu Gln Lys Asn Thr Thr Ile Asp Val Arg		
	110	120
Thr Thr Phe Ser Arg Cys Leu Ser Lys Glu Trp Gly Ser Phe Leu		
	125	135
Glu Glu Ser Tyr Pro Tyr Phe Leu Ile Val Ala Asp Glu Gly Leu		
	140	150
Asn Asp Leu Gln Thr Gln Leu Phe Asn Phe Leu Ile Ile His Ser		
	155	165
Trp Ala Arg Lys Val Asn Val Val Leu Ser Ser Gly Gln Glu Ser		
	170	180
Asp Val Leu Cys Leu Tyr Ala Tyr Leu Leu Pro Ser Met Tyr Arg		

	185		190		195
His Gln Ile Phe Ser Trp Lys Asn Lys Gln Asn Ile Lys Asp Ala					
	200		205		210
Tyr Thr Thr Leu Leu Asn Gln Leu Glu Arg Phe Lys Leu Ser Ala					
	215		220		225
Leu Ala Pro Leu Phe Gly Ser Leu Lys Trp Asn Asn Ile Thr Glu					
	230		235		240
Glu Ala His Lys Thr Val Ser Leu Leu Thr Gln Val Trp Pro Glu					
	245		250		255
Gly Ser Asp Ile Arg Arg Val Phe Cys Val Thr Ser Cys Ser Leu					
	260		265		270
Ser Leu Arg Met Tyr His Arg Phe Leu Gly Asn Arg Glu Pro Ser					
	275		280		285
Ser Gly Gln Glu Thr Glu Ile Gln Gln Val Asn Ser Asn Cys Leu					
	290		295		300
Thr Leu Gln Glu Met Glu Asp Leu Cys Lys Leu His Cys Leu Thr					
	305		310		315
Val Val Phe Leu Leu His Leu Pro Leu Ser Gln Arg Ala Cys Ala					
	320		325		330
Arg Val Ile Thr Ser His Trp Ala Glu Asp Met Lys Pro Leu Leu					
	335		340		345
Gln Met Lys Lys Trp Cys Glu Tyr Phe Ile Leu Arg Asn Ile His					
	350		355		360
Thr Phe Glu Phe Trp Asn Leu Asn Leu Ile His Leu Ser Asp Leu					
	365		370		375
Asn Asp Glu Leu Leu Leu Lys Asn Ile Ala Phe Tyr Tyr Glu Asn					
	380		385		390
Glu Asn Val Lys Gly Leu His Leu Asn Leu Gly Asp Thr Ile Met					
	395		400		405
Lys Asp Tyr Glu Tyr Leu Trp Asn Thr Ile Ser Lys Leu Val Arg					
	410		415		420
Asp Phe Glu Val Gly Gln Pro Phe Pro Leu Arg Thr Thr Lys Val					
	425		430		435
Cys Phe Leu Glu Lys Lys Pro Ser Pro Ile Lys Asp Ser Ser Asn					
	440		445		450
Glu Met Val Pro Asn Leu Gly Phe Ile Pro Thr Ser Ser Phe Val					
	455		460		465
Val Asp Lys Phe Ala Gly Asp Ile Leu Lys Asp Leu Pro Phe Leu					
	470		475		480
Lys Ser Asp Asp Pro Ile Val Thr Ser Leu Val Lys Gln Lys Glu					
	485		490		495
Phe Asp Glu Leu Val His Trp His Ser His Lys Pro Leu Ser Asp					
	500		505		510
Asp Tyr Asp Arg Ser Arg Cys Gln Phe Asp Glu Lys Ser Arg Asp					
	515		520		525
Pro Arg Val Leu Arg Ser Val Gln Lys Tyr His Val Phe Gln Arg					
	530		535		540
Phe Tyr Gly Asn Ser Leu Glu Thr Val Ser Ser Lys Ile Ile Val					
	545		550		555
Thr Gln Thr Ile Lys Ser Lys Lys Asp Phe Ser Gly Pro Lys Ser					
	560		565		570
Lys Lys Ala His Glu Thr Lys Ala Glu Ile Ile Ala Arg Glu Asn					
	575		580		585
Lys Lys Arg Leu Phe Ala Arg Glu Glu Gln Lys Glu Glu Gln Lys					
	590		595		600
Trp Asn Ala Leu Ser Phe Ser Ile Glu Glu Gln Leu Lys Glu Asn					

	605	610	615
Leu His Ser Gly	Ile Lys Ser Leu Glu Asp Phe Leu Lys Ser Cys		
	620	625	630
Lys Ser Ser Cys	Val Lys Leu Gln Val Glu Met Val Gly Leu Thr		
	635	640	645
Ala Cys Leu Lys	Ala Trp Lys Glu His Cys Arg Ser Glu Glu Gly		
	650	655	660
Lys Thr Thr Lys	Asp Leu Ser Ile Ala Val Gln Val Met Lys Arg		
	665	670	675
Ile His Ser Leu	Met Glu Lys Tyr Ser Glu Leu Leu Gln Glu Asp		
	680	685	690
Asp Arg Gln Leu	Ile Ala Arg Cys Leu Lys Tyr Leu Gly Phe Asp		
	695	700	705
Glu Leu Ala Ser	Ser Leu His Pro Ala Gln Asp Ala Glu Asn Asp		
	710	715	720
Val Lys Val Lys	Lys Arg Asn Lys Tyr Ser Val Gly Ile Gly Pro		
	725	730	735
Ala Arg Phe Gln	Leu Gln Tyr Met Gly His Tyr Leu Ile Arg Asp		
	740	745	750
Glu Arg Lys Asp	Pro Asp Pro Arg Val Gln Asp Phe Ile Pro Asp		
	755	760	765
Thr Trp Gln Arg	Glu Leu Leu Asp Val Val Asp Lys Asn Glu Ser		
	770	775	780
Ala Val Ile Val	Ala Pro Thr Ser Ser Gly Lys Thr Tyr Ala Ser		
	785	790	795
Tyr Tyr Cys Met	Glu Lys Val Leu Lys Glu Ser Asp Asp Gly Val		
	800	805	810
Val Val Tyr Val	Ala Pro Thr Lys Ala Leu Val Asn Gln Val Ala		
	815	820	825
Ala Thr Val Gln	Asn Arg Phe Thr Lys Asn Leu Pro Ser Gly Glu		
	830	835	840
Val Leu Cys Gly	Val Phe Thr Arg Glu Tyr Arg His Asp Ala Leu		
	845	850	855
Asn Cys Gln Val	Leu Ile Thr Val Pro Ala Cys Phe Glu Ile Leu		
	860	865	870
Leu Leu Ala Pro	His Arg Gln Asn Trp Val Lys Lys Ile Arg Tyr		
	875	880	885
Val Ile Phe Asp	Glu Val His Cys Leu Gly Gly Glu Ile Gly Ala		
	890	895	900
Glu Ile Trp Glu	His Leu Leu Val Met Ile Arg Cys Pro Phe Leu		
	905	910	915
Ala Leu Ser Ala	Thr Ile Ser Asn Pro Glu His Leu Thr Glu Trp		
	920	925	930
Leu Gln Ser Val	Lys Trp Tyr Trp Lys Gln Glu Asp Lys Ile Ile		
	935	940	945
Glu Asn Asn Thr	Ala Ser Lys Arg His Val Gly Arg Gln Ala Gly		
	950	955	960
Phe Pro Lys Asp	Tyr Leu Gln Val Lys Gln Ser Tyr Lys Val Arg		
	965	970	975
Leu Val Leu Tyr	Gly Glu Arg Tyr Asn Asp Leu Glu Lys His Val		
	980	985	990
Cys Ser Ile Lys	His Gly Asp Ile His Phe Asp His Phe His Pro		
	995	1000	1005
Cys Ala Ala Leu	Thr Thr Asp His Ile Glu Arg Tyr Gly Phe Pro		
	1010	1015	1020
Pro Asp Leu Thr	Leu Ser Pro Arg Glu Ser Ile Gln Leu Tyr Asp		

1025	1030	1035
Ala Met Phe Gln Ile Trp Lys Ser Trp Pro Arg Ala Gln Glu Leu		
1040	1045	1050
Cys Pro Glu Asn Phe Ile His Phe Asn Asn Lys Leu Val Ile Lys		
1055	1060	1065
Lys Met Asp Ala Arg Lys Tyr Glu Glu Ser Leu Lys Ala Glu Leu		
1070	1075	1080
Thr Ser Trp Ile Lys Asn Gly Asn Val Glu Gln Ala Arg Met Val		
1085	1090	1095
Leu Gln Asn Leu Ser Pro Glu Ala Asp Leu Ser Pro Glu Asn Met		
1100	1105	1110
Ile Thr Met Phe Pro Leu Leu Val Glu Lys Leu Arg Lys Met Glu		
1115	1120	1125
Lys Leu Pro Ala Leu Phe Phe Leu Phe Lys Leu Gly Ala Val Glu		
1130	1135	1140
Asn Ala Ala Glu Ser Val Ser Thr Phe Leu Lys Lys Lys Gln Glu		
1145	1150	1155
Thr Lys Arg Pro Pro Lys Ala Asp Lys Glu Ala His Val Met Ala		
1160	1165	1170
Asn Lys Leu Arg Lys Val Lys Lys Ser Ile Glu Lys Gln Lys Ile		
1175	1180	1185
Ile Asp Glu Lys Ser Gln Lys Lys Thr Arg Asn Val Asp Gln Ser		
1190	1195	1200
Leu Ile His Glu Ala Glu His Asp Asn Leu Val Lys Cys Leu Glu		
1205	1210	1215
Lys Asn Leu Glu Ile Pro Gln Asp Cys Thr Tyr Ala Asp Gln Lys		
1220	1225	1230
Ala Val Asp Thr Glu Thr Leu Gln Arg Val Phe Gly Arg Val Lys		
1235	1240	1245
Phe Glu Arg Lys Gly Glu Glu Leu Lys Ala Leu Ala Glu Arg Gly		
1250	1255	1260
Ile Gly Tyr His His Ser Ala Met Ser Phe Lys Glu Lys Gln Leu		
1265	1270	1275
Val Glu Ile Leu Phe Arg Lys Gly Tyr Leu Arg Val Val Thr Ala		
1280	1285	1290
Thr Gly Thr Leu Ala Leu Gly Val Asn Met Pro Cys Lys Ser Val		
1295	1300	1305
Val Phe Ala Gln Asn Ser Val Tyr Leu Asp Ala Leu Asn Tyr Arg		
1310	1315	1320
Gln Met Ser Gly Arg Ala Gly Arg Arg Gly Gln Asp Leu Met Gly		
1325	1330	1335
Asp Val Tyr Phe Phe Asp Ile Pro Phe Pro Lys Ile Gly Lys Leu		
1340	1345	1350
Ile Lys Ser Asn Val Pro Glu Leu Arg Gly His Phe Pro Leu Ser		
1355	1360	1365
Ile Thr Leu Val Leu Arg Leu Met Leu Leu Ala Ser Lys Gly Asp		
1370	1375	1380
Asp Pro Glu Asp Ala Lys Ala Lys Val Leu Ser Val Leu Lys His		
1385	1390	1395
Ser Leu Leu Ser Phe Lys Gln Pro Arg Val Met Asp Met Leu Lys		
1400	1405	1410
Leu Tyr Phe Leu Phe Ser Leu Gln Phe Leu Val Lys Glu Gly Tyr		
1415	1420	1425
Leu Asp Gln Glu Gly Asn Pro Met Gly Phe Ala Gly Leu Val Ser		
1430	1435	1440
His Leu His Tyr His Glu Pro Ser Asn Leu Val Phe Val Ser Phe		

1445	1450	1455
Leu Val Asn Gly Leu Phe His Asp	Leu Cys Gln Pro Thr Arg Lys	
1460	1465	1470
Gly Ser Lys His Phe Ser Gln Asp	Val Met Glu Lys Leu Val Leu	
1475	1480	1485
Val Leu Ala His Leu Phe Gly Arg	Arg Tyr Phe Pro Pro Lys Phe	
1490	1495	1500
Gln Asp Ala His Phe Glu Phe Tyr	Gln Ser Lys Val Phe Leu Asp	
1505	1510	1515
Asp Leu Pro Glu Asp Phe Ser Asp	Ala Leu Asp Glu Tyr Asn Met	
1520	1525	1530
Lys Ile Met Glu Asp Phe Thr Thr	Phe Leu Arg Ile Val Ser Lys	
1535	1540	1545
Leu Ala Asp Met Asn Gln Glu Tyr	Gln Leu Pro Leu Ser Lys Ile	
1550	1555	1560
Lys Phe Thr Gly Lys Glu Cys Glu	Asp Ser Gln Leu Val Ser His	
1565	1570	1575
Leu Met Ser Cys Lys Glu Gly Arg	Val Ala Ile Ser Pro Phe Val	
1580	1585	1590
Cys Leu Ser Gly Asn Phe Asp Asp	Asp Leu Leu Arg Leu Glu Thr	
1595	1600	1605
Pro Asn His Val Thr Leu Gly Thr	Ile Gly Val Asn Arg Ser Gln	
1610	1615	1620
Ala Pro Val Leu Leu Ser Gln Lys	Phe Asp Asn Arg Gly Arg Lys	
1625	1630	1635
Met Ser Leu Asn Ala Tyr Ala Leu	Asp Phe Tyr Lys His Gly Ser	
1640	1645	1650
Leu Ile Gly Leu Val Gln Asp Asn	Arg Met Asn Glu Gly Asp Ala	
1655	1660	1665
Tyr Tyr Leu Leu Lys Asp Phe	Ala Leu Thr Ile Lys Ser Ile Ser	
1670	1675	1680
Val Ser Leu Arg Glu Leu Cys Glu	Asn Glu Asp Asp Asn Val Val	
1685	1690	1695
Leu Ala Phe Glu Gln Leu Ser Thr	Thr Phe Trp Glu Lys Leu Asn	
1700	1705	1710
Lys Val		

&lt;210&gt; 31

&lt;211&gt; 780

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7473738CD1

&lt;400&gt; 31

Met Ser Arg Phe Pro Ala Val Ala Gly Arg Ala Pro Arg Arg Gln	
1 5 10 15	
Glu Glu Gly Glu Arg Pro Ile Glu Leu Gln Glu Glu Arg Pro Ser	
20 25 30	
Ala Val Arg Ile Ala Asp Arg Glu Glu Lys Gly Cys Thr Ser Gln	
35 40 45	
Glu Gly Gly Thr Thr Pro Thr Phe Pro Ile Gln Lys Gln Arg Lys	
50 55 60	

Lys	Leu	Ile	Gln	Ala	Val	Arg	Asp	Asn	Ser	Phe	Leu	Ile	Val	Thr	65	70	75
Gly	Asn	Thr	Gly	Ser	Gly	Lys	Thr	Thr	Gln	Leu	Pro	Lys	Tyr	Leu	80	85	90
Tyr	Glu	Ala	Gly	Phe	Ser	Gln	His	Gly	Met	Ile	Gly	Val	Thr	Gln	95	100	105
Pro	Arg	Lys	Val	Ala	Ala	Ile	Ser	Val	Ala	Gln	Arg	Val	Ala	Glu	110	115	120
Glu	Met	Lys	Cys	Thr	Leu	Gly	Ser	Lys	Val	Gly	Tyr	Gln	Val	Arg	125	130	135
Phe	Asp	Asp	Cys	Ser	Ser	Lys	Glu	Thr	Ala	Ile	Lys	Tyr	Met	Thr	140	145	150
Asp	Gly	Cys	Leu	Leu	Lys	His	Ile	Leu	Gly	Asp	Pro	Asn	Leu	Thr	155	160	165
Lys	Phe	Ser	Val	Ile	Ile	Leu	Asp	Glu	Ala	His	Glu	Arg	Thr	Leu	170	175	180
Thr	Thr	Asp	Ile	Leu	Phe	Gly	Leu	Leu	Lys	Lys	Leu	Phe	Gln	Glu	185	190	195
Lys	Ser	Pro	Asn	Arg	Lys	Glu	His	Leu	Thr	Ser	Gly	Gly	Thr	Cys	200	205	210
His	Ala	Thr	Met	Glu	Leu	Ala	Lys	Leu	Ser	Ala	Phe	Phe	Gly	Asn	215	220	225
Cys	Pro	Ile	Phe	Asp	Ile	Pro	Gly	Arg	Leu	Tyr	Pro	Val	Arg	Glu	230	235	240
Lys	Phe	Cys	Asn	Leu	Ile	Gly	Pro	Arg	Asp	Arg	Glu	Asn	Thr	Ala	245	250	255
Tyr	Ile	Gln	Ala	Ile	Val	Lys	Val	Thr	Met	Asp	Ile	His	Leu	Asn	260	265	270
Glu	Met	Ala	Gly	Asp	Ile	Leu	Val	Phe	Leu	Thr	Gly	Gln	Phe	Glu	275	280	285
Ile	Glu	Lys	Ser	Cys	Glu	Leu	Leu	Phe	Gln	Met	Ala	Glu	Ser	Val	290	295	300
Asp	Tyr	Asp	Tyr	Asp	Val	Gln	Asp	Thr	Thr	Leu	Asp	Gly	Leu	Leu	305	310	315
Ile	Leu	Pro	Cys	Tyr	Gly	Ser	Met	Thr	Thr	Asp	Gln	Gln	Arg	Arg	320	325	330
Ile	Phe	Leu	Pro	Pro	Pro	Pro	Gly	Ile	Arg	Lys	Cys	Val	Ile	Ser	335	340	345
Thr	Asn	Ile	Ser	Ala	Thr	Ser	Leu	Thr	Ile	Asp	Gly	Ile	Arg	Tyr	350	355	360
Val	Val	Asp	Gly	Gly	Phe	Val	Lys	Gln	Leu	Asn	His	Asn	Pro	Arg	365	370	375
Leu	Gly	Leu	Asp	Ile	Leu	Glu	Val	Val	Pro	Ile	Ser	Lys	Ser	Glu	380	385	390
Ala	Leu	Gln	Arg	Ser	Gly	Arg	Ala	Gly	Arg	Thr	Ser	Ser	Gly	Lys	395	400	405
Cys	Phe	Arg	Ile	Tyr	Ser	Lys	Asp	Phe	Trp	Asn	Gln	Cys	Met	Pro	410	415	420
Asp	His	Val	Ile	Pro	Glu	Ile	Lys	Arg	Thr	Ser	Leu	Thr	Ser	Val	425	430	435
Val	Leu	Thr	Leu	Lys	Cys	Leu	Ala	Ile	His	Asp	Val	Ile	Arg	Phe	440	445	450
Pro	Tyr	Leu	Asp	Pro	Pro	Asn	Glu	Arg	Leu	Ile	Leu	Glu	Ala	Leu	455	460	465
Lys	Gln	Leu	Tyr	Gln	Cys	Asp	Ala	Ile	Asp	Arg	Ser	Gly	His	Val	470	475	480

Thr	Arg	Leu	Gly	Leu	Ser	Met	Val	Glu	Phe	Pro	Leu	Pro	Pro	His
				485					490					495
Leu	Thr	Cys	Ala	Val	Ile	Lys	Ala	Ala	Ser	Leu	Asp	Cys	Glu	Asp
				500					505					510
Leu	Leu	Leu	Pro	Ile	Ala	Ala	Met	Leu	Ser	Val	Glu	Asn	Val	Phe
				515					520					525
Ile	Arg	Pro	Val	Asp	Pro	Glu	Tyr	Gln	Lys	Glu	Ala	Glu	Gln	Arg
				530					535					540
His	Arg	Glu	Leu	Ala	Ala	Lys	Ala	Gly	Gly	Phe	Asn	Asp	Phe	Ala
				545					550					555
Thr	Leu	Ala	Val	Ile	Phe	Glu	Gln	Cys	Lys	Ser	Ser	Gly	Ala	Pro
				560					565					570
Ala	Ser	Trp	Cys	Gln	Lys	His	Trp	Ile	His	Trp	Arg	Cys	Leu	Phe
				575					580					585
Ser	Ala	Phe	Arg	Val	Glu	Ala	Gln	Leu	Arg	Glu	Leu	Ile	Arg	Lys
				590					595					600
Leu	Lys	Gln	Gln	Ser	Asp	Phe	Pro	Lys	Glu	Thr	Phe	Glu	Gly	Pro
				605					610					615
Lys	His	Glu	Val	Leu	Arg	Arg	Cys	Leu	Cys	Ala	Gly	Tyr	Phe	Lys
				620					625					630
Asn	Val	Ala	Arg	Arg	Ser	Val	Gly	Arg	Thr	Phe	Cys	Thr	Met	Asp
				635					640					645
Gly	Arg	Gly	Ser	Pro	Val	His	Ile	His	Pro	Ser	Ser	Ala	Leu	His
				650					655					660
Glu	Gln	Glu	Thr	Lys	Leu	Glu	Trp	Ile	Ile	Phe	His	Glu	Val	Leu
				665					670					675
Val	Thr	Thr	Lys	Val	Tyr	Ala	Arg	Ile	Val	Cys	Pro	Ile	Arg	Tyr
				680					685					690
Glu	Trp	Val	Arg	Asp	Leu	Leu	Pro	Lys	Leu	His	Glu	Leu	Asn	Ala
				695					700					705
His	Asp	Leu	Ser	Ser	Val	Ala	Arg	Arg	Glu	Met	Arg	Glu	Asp	Ala
				710					715					720
Arg	Arg	Lys	Trp	Thr	Asn	Lys	Glu	Asn	Val	Lys	Gln	Leu	Lys	Asp
				725					730					735
Gly	Ile	Ser	Lys	Glu	Val	Leu	Lys	Lys	Met	Gln	Arg	Arg	Asn	Asp
				740					745					750
Asp	Lys	Ser	Ile	Ser	Asp	Ala	Arg	Ala	Arg	Phe	Leu	Glu	Arg	Lys
				755					760					765
Gln	Gln	Arg	Ile	Gln	Asp	His	Ser	Asp	Thr	Leu	Lys	Glu	Thr	Gly
				770					775					780

&lt;210&gt; 32

&lt;211&gt; 648

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4447743CD1

&lt;400&gt; 32

Met	Glu	Leu	Val	Thr	Phe	Arg	Asp	Val	Ala	Ile	Glu	Phe	Ser	Pro
1				5					10					15
Glu	Glu	Trp	Lys	Cys	Leu	Asp	Pro	Ala	Gln	Gln	Asn	Leu	Tyr	Arg
				20					25					30

Asp Val Met Leu Glu Asn Tyr Arg Asn Leu Val Ser Leu Gly Phe		
	35	40 45
Val Ile Ser Asn Pro Asp Leu Val Thr Cys Leu Glu Gln Ile Lys		
	50	55 60
Glu Pro Cys Asn Leu Lys Ile His Glu Thr Ala Ala Lys Pro Pro		
	65	70 75
Ala Ile Cys Ser Pro Phe Ser Gln Asp Leu Ser Pro Val Gln Gly		
	80	85 90
Ile Glu Asp Ser Phe His Lys Leu Ile Leu Lys Arg Tyr Glu Lys		
	95	100 105
Cys Gly His Glu Asn Leu Gln Leu Arg Lys Gly Cys Lys Arg Val		
	110	115 120
Asn Glu Cys Lys Val Gln Lys Gly Val Asn Asn Gly Val Tyr Gln		
	125	130 135
Cys Leu Ser Thr Thr Gln Ser Lys Ile Phe Gln Cys Asn Thr Cys		
	140	145 150
Val Lys Val Phe Ser Lys Phe Ser Asn Ser Asn Lys His Lys Ile		
	155	160 165
Arg His Thr Gly Glu Lys Pro Phe Lys Cys Thr Glu Cys Gly Arg		
	170	175 180
Ser Phe Tyr Met Ser His Leu Thr Gln His Thr Gly Ile His Ala		
	185	190 195
Gly Glu Lys Pro Tyr Lys Cys Glu Lys Cys Gly Lys Ala Phe Asn		
	200	205 210
Arg Ser Thr Ser Leu Ser Lys His Lys Arg Ile His Thr Gly Glu		
	215	220 225
Lys Pro Tyr Thr Cys Glu Glu Cys Gly Lys Ala Phe Arg Arg Ser		
	230	235 240
Thr Val Leu Asn Glu His Lys Lys Ile His Thr Gly Glu Lys Pro		
	245	250 255
Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Thr Arg Ser Thr Thr		
	260	265 270
Leu Asn Glu His Lys Lys Ile His Thr Gly Glu Lys Pro Tyr Lys		
	275	280 285
Cys Lys Glu Cys Gly Lys Ala Phe Arg Trp Ser Thr Ser Leu Asn		
	290	295 300
Glu His Lys Asn Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Lys		
	305	310 315
Glu Cys Gly Lys Ala Phe Arg Gln Ser Arg Ser Leu Asn Glu His		
	320	325 330
Lys Asn Ile His Thr Gly Glu Lys Pro Tyr Thr Cys Glu Lys Cys		
	335	340 345
Gly Lys Ala Phe Asn Gln Ser Ser Ser Leu Ile Ile His Arg Ser		
	350	355 360
Ile His Ser Glu Gln Lys Leu Tyr Lys Cys Glu Glu Cys Gly Lys		
	365	370 375
Ala Phe Thr Trp Ser Ser Ser Leu Asn Lys His Lys Arg Ile His		
	380	385 390
Thr Gly Glu Lys Pro Tyr Thr Cys Glu Glu Cys Gly Lys Ala Phe		
	395	400 405
Tyr Arg Ser Ser His Leu Ala Lys His Lys Arg Ile His Thr Gly		
	410	415 420
Glu Lys Pro Tyr Thr Cys Glu Glu Cys Gly Lys Ala Phe Asn Gln		
	425	430 435
Ser Ser Thr Leu Ile Leu His Lys Arg Ile His Ser Gly Gln Lys		
	440	445 450

```

Pro Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Thr Arg Ser Thr
      455                      460                      465
Thr Leu Asn Glu His Lys Lys Ile His Thr Gly Glu Lys Pro Tyr
      470                      475                      480
Lys Cys Glu Glu Cys Gly Lys Ala Phe Ile Trp Ser Ala Ser Leu
      485                      490                      495
Asn Glu His Lys Asn Ile His Thr Gly Glu Lys Pro Tyr Lys Cys
      500                      505                      510
Lys Glu Cys Gly Lys Ala Phe Asn Gln Ser Ser Gly Leu Ile Ile
      515                      520                      525
His Arg Ser Ile His Ser Glu Gln Lys Leu Tyr Lys Cys Glu Glu
      530                      535                      540
Cys Gly Lys Ala Phe Thr Arg Ser Thr Ala Leu Asn Glu His Lys
      545                      550                      555
Lys Ile His Ser Gly Glu Lys Pro Tyr Lys Cys Lys Glu Cys Gly
      560                      565                      570
Lys Ala Tyr Asn Leu Ser Ser Thr Leu Thr Lys His Lys Arg Ile
      575                      580                      585
His Thr Gly Glu Lys Pro Phe Thr Cys Glu Glu Cys Gly Lys Ala
      590                      595                      600
Phe Asn Trp Ser Ser Ser Leu Thr Lys His Lys Ile Ile His Thr
      605                      610                      615
Gly Glu Lys Ser Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Asn
      620                      625                      630
Arg Pro Ser Thr Leu Thr Val His Lys Arg Ile His Thr Gly Lys
      635                      640                      645
Glu His Ser

```

&lt;210&gt; 33

&lt;211&gt; 602

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7497554CD1

&lt;400&gt; 33

```

Met Ser Glu Arg Arg Arg Ser Ala Val Ala Leu Ser Ser Arg Ala
  1                      5                      10                      15
His Ala Phe Ser Val Glu Ala Leu Ile Gly Ser Asn Lys Lys Arg
      20                      25                      30
Lys Leu Arg Asp Trp Glu Glu Lys Gly Leu Asp Leu Ser Met Glu
      35                      40                      45
Ala Leu Ser Pro Ala Gly Pro Leu Gly Asp Thr Glu Asp Ala Ala
      50                      55                      60
Ala His Gly Leu Glu Pro His Pro Asp Ser Glu Gln Ser Thr Gly
      65                      70                      75
Ser Asp Ser Glu Val Leu Thr Glu Arg Thr Ser Cys Ser Phe Ser
      80                      85                      90
Thr His Thr Asp Leu Ala Ser Gly Ala Ala Gly Pro Val Pro Ala
      95                      100                      105
Ala Met Ser Ser Met Glu Glu Ile Gln Val Glu Leu Gln Cys Ala
      110                      115                      120
Asp Leu Trp Lys Arg Phe His Asp Ile Gly Thr Glu Met Ile Ile

```

	125		130		135
Thr Lys Ala Gly Arg Arg Met Phe Pro Ala Met Arg Val Lys Ile					
	140		145		150
Thr Gly Leu Asp Pro Asn Gln Gln Tyr Tyr Ile Ala Met Asp Ile					
	155		160		165
Val Pro Val Asp Asn Lys Arg Tyr Arg Tyr Val Tyr His Ser Ser					
	170		175		180
Lys Trp Met Val Ala Gly Asn Ala Asp Ser Pro Val Pro Pro Arg					
	185		190		195
Val Tyr Ile His Pro Asp Ser Leu Ala Ser Gly Asp Thr Trp Met					
	200		205		210
Arg Gln Val Val Ser Phe Asp Lys Leu Lys Leu Thr Asn Asn Glu					
	215		220		225
Leu Asp Asp Gln Gly His Ile Ile Leu His Ser Met His Lys Tyr					
	230		235		240
Gln Pro Arg Val His Val Ile Arg Lys Asp Phe Ser Ser Asp Leu					
	245		250		255
Ser Pro Thr Lys Pro Val Pro Val Gly Asp Gly Val Lys Thr Phe					
	260		265		270
Asn Phe Pro Glu Thr Val Phe Thr Thr Val Thr Ala Tyr Gln Asn					
	275		280		285
Gln Gln Ile Thr Arg Leu Lys Ile Asp Arg Asn Pro Phe Ala Lys					
	290		295		300
Gly Phe Arg Asp Ser Gly Arg Asn Arg Thr Gly Leu Glu Ala Ile					
	305		310		315
Met Glu Thr Tyr Ala Phe Trp Arg Pro Pro Val Arg Thr Leu Thr					
	320		325		330
Phe Glu Asp Phe Thr Thr Met Gln Lys Gln Gln Gly Gly Ser Thr					
	335		340		345
Gly Thr Ser Pro Thr Thr Ser Ser Thr Gly Thr Pro Ser Pro Ser					
	350		355		360
Ala Ser Ser His Leu Leu Ser Pro Ser Cys Ser Pro Pro Thr Phe					
	365		370		375
His Leu Ala Pro Asn Thr Phe Asn Val Gly Cys Arg Glu Ser Gln					
	380		385		390
Leu Cys Asn Leu Asn Leu Ser Asp Tyr Pro Pro Cys Ala Arg Ser					
	395		400		405
Asn Met Ala Ala Leu Gln Ser Tyr Pro Gly Leu Ser Asp Ser Gly					
	410		415		420
Tyr Asn Arg Leu Gln Ser Gly Thr Thr Ser Ala Thr Gln Pro Ser					
	425		430		435
Glu Thr Phe Met Pro Gln Arg Thr Pro Ser Leu Ile Ser Gly Ile					
	440		445		450
Pro Thr Pro Pro Ser Leu Pro Gly Asn Ser Lys Met Glu Ala Tyr					
	455		460		465
Gly Gly Gln Leu Gly Ser Phe Pro Thr Ser Gln Phe Gln Tyr Val					
	470		475		480
Met Gln Ala Gly Asn Ala Ala Ser Ser Ser Ser Ser Pro His Met					
	485		490		495
Phe Gly Gly Ser His Met Gln Gln Ser Ser Tyr Asn Ala Phe Ser					
	500		505		510
Leu His Asn Pro Tyr Asn Leu Tyr Gly Tyr Asn Phe Pro Thr Ser					
	515		520		525
Pro Arg Leu Ala Ala Ser Pro Glu Lys Leu Ser Ala Ser Gln Ser					
	530		535		540
Thr Leu Leu Cys Ser Ser Pro Ser Asn Gly Ala Phe Gly Glu Arg					

	545		550		555
Gln Tyr Leu Pro	Ser Gly Met Glu His	Ser Met His Met Ile	Ser		
	560		565		570
Pro Ser Pro Asn	Asn Gln Gln Ala Thr	Asn Thr Cys Asp Gly	Arg		
	575		580		585
Gln Tyr Gly Ala	Val Pro Gly Ser Ser	Ser Gln Met Ser Val	His		
	590		595		600
Met Val					

&lt;210&gt; 34

&lt;211&gt; 388

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7475843CD1

&lt;400&gt; 34

Met Leu Glu Asn Tyr	Arg Asn Leu Val	Ser Leu Gly Ile Ala	Val
1	5	10	15
Ser Lys Pro Asp Leu	Ile Thr Cys Leu	Glu Gln Arg Asn Glu	Pro
	20	25	30
Trp Asn Val Lys Lys	His Glu Thr Val	Ala Arg His Pro Ala	Val
	35	40	45
Ser Ser His Phe Thr	Gln Asp Leu Leu	Pro Glu His Gly Ile	Lys
	50	55	60
Asp Ser Phe Gln Lys	Val Ile Leu Arg	Arg Tyr Gly Ser Tyr	Gly
	65	70	75
Ile Glu Asn Leu Gln	Leu Lys Lys Asp	Trp Glu Ser Val Gly	Glu
	80	85	90
Ser Lys Val Gln Lys	Glu Cys Cys Asn	Gly Leu Asn Gln Ser	Leu
	95	100	105
Ser Thr Thr His Thr	Lys Ile Phe Gln	Phe Asn Lys Cys Val	Lys
	110	115	120
Val Phe Ser Lys Ser	Ser Asn Leu Asn	Arg His Lys Ile Arg	His
	125	130	135
Thr Gly Glu Ile Ser	Ser Asn Cys Lys	Glu Cys Asp Asn Ser	Phe
	140	145	150
Tyr Ile Ser Ser Val	Leu Thr Pro Leu	Gln Arg Ile His Thr	Ala
	155	160	165
Glu Lys Ser Tyr Lys	Cys Lys Gln Cys	Gly Lys Ala Phe Arg	His
	170	175	180
Cys Ser Cys Phe Leu	Glu His Glu Thr	Ile His Asn Glu Glu	Lys
	185	190	195
His Tyr Lys Cys Lys	Glu Cys Gly Lys	Val Phe Lys Ser Phe	Thr
	200	205	210
Ser Leu Ser Asn His	Ile Ile Ile His	Thr Gly Lys Lys Leu	Tyr
	215	220	225
Lys Cys Glu Glu Cys	Gly Lys Ala Phe	Asn His Ser Ser Asn	His
	230	235	240
Ala Lys His Lys Lys	Ile His Thr Gly	Gln Lys Pro His Lys	Cys
	245	250	255
Glu Glu Cys Gly Lys	Ala Phe Asn Trp	Phe Ser Tyr Leu Thr	Leu
	260	265	270

His Lys Arg Ile	His Thr Gly Glu Lys	Pro Tyr Lys Cys Asp	Glu
275	280		285
Cys Gly Lys Ala	Phe Asn Gln Cys Ser	Asn Leu Thr Lys His	Lys
290	295		300
Arg Ile His Thr	Gly Glu Lys Pro Tyr	Lys Cys Glu Glu Cys	Gly
305	310		315
Lys Ala Phe Asn	Arg Cys Ser His Leu	Thr Glu His Lys Arg	Ile
320	325		330
His Thr Gly Glu	Lys Pro Tyr Lys Cys	Glu Glu Cys Gly Lys	Val
335	340		345
Phe Ile Ser Cys	Ser Ser Leu Ser Asn	His Lys Arg Ile His	Thr
350	355		360
Arg Glu Lys Cys	Tyr Lys Ser Glu Glu	Cys Gly Lys Thr Phe	Asn
365	370		375
His Cys Ser Asp	Leu Asn Val Pro Glu	Lys Ile His Thr	
380	385		

&lt;210&gt; 35

&lt;211&gt; 480

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6319550CD1

&lt;400&gt; 35

Met Gly Ser Pro	Ala Ala Pro Glu Gly	Ala Leu Gly Tyr Val	Arg
1	5	10	15
Glu Phe Thr Arg	His Ser Ser Asp Val	Leu Gly Asn Leu Asn	Glu
20	25		30
Leu Arg Leu Arg	Gly Ile Leu Thr Asp	Val Thr Leu Leu Val	Gly
35	40		45
Gly Gln Pro Leu	Arg Ala His Lys Ala	Val Leu Ile Ala Cys	Ser
50	55		60
Gly Phe Phe Tyr	Ser Ile Phe Arg Gly	Arg Ala Gly Val Gly	Val
65	70		75
Asp Val Leu Ser	Leu Pro Gly Gly Pro	Glu Ala Arg Gly Phe	Ala
80	85		90
Pro Leu Leu Asp	Phe Met Tyr Thr Ser	Arg Leu Arg Leu Ser	Pro
95	100		105
Ala Thr Ala Pro	Ala Val Leu Ala Ala	Ala Thr Tyr Leu Gln	Met
110	115		120
Glu His Val Val	Gln Ala Cys His Arg	Phe Ile Gln Ala Ser	Tyr
125	130		135
Glu Pro Leu Gly	Ile Ser Leu Arg Pro	Leu Glu Ala Glu Pro	Pro
140	145		150
Thr Pro Pro Thr	Ala Pro Pro Pro Gly	Ser Pro Arg Arg Ser	Glu
155	160		165
Gly His Pro Asp	Pro Pro Thr Glu Ser	Arg Ser Cys Ser Gln	Gly
170	175		180
Pro Pro Ser Pro	Ala Ser Pro Asp Pro	Lys Ala Cys Asn Trp	Lys
185	190		195
Lys Tyr Lys Tyr	Ile Val Leu Asn Ser	Gln Ala Ser Gln Ala	Gly
200	205		210
Ser Leu Val Gly	Glu Arg Ser Ser Gly	Gln Pro Cys Pro Gln	Ala

	215		220		225
Arg Leu Pro Ser	Gly Asp Glu Ala Ser	Ser Ser Ser Ser Ser Ser			
	230		235		240
Ser Ser Ser Ser	Ser Glu Glu Gly Pro	Ile Pro Gly Pro Gln Ser			
	245		250		255
Arg Leu Ser Pro	Thr Ala Ala Thr Val	Gln Phe Lys Cys Gly Ala			
	260		265		270
Pro Ala Ser Thr	Pro Tyr Leu Leu Thr	Ser Gln Ala Gln Asp Thr			
	275		280		285
Ser Gly Ser Pro	Ser Glu Arg Ala Arg	Pro Leu Pro Gly Ser Glu			
	290		295		300
Phe Phe Ser Cys	Gln Asn Cys Glu Ala	Val Ala Gly Cys Ser Ser			
	305		310		315
Gly Leu Asp Ser	Leu Val Pro Gly Asp	Glu Asp Lys Pro Tyr Lys			
	320		325		330
Cys Gln Leu Cys	Arg Ser Ser Phe Arg	Tyr Lys Gly Asn Leu Ala			
	335		340		345
Ser His Arg Thr	Val His Thr Gly Glu	Lys Pro Tyr His Cys Ser			
	350		355		360
Ile Cys Gly Ala	Arg Phe Asn Arg Pro	Ala Asn Leu Lys Thr His			
	365		370		375
Ser Arg Ile His	Ser Gly Glu Lys Pro	Tyr Lys Cys Glu Thr Cys			
	380		385		390
Gly Ser Arg Phe	Val Gln Val Ala His	Leu Arg Ala His Val Leu			
	395		400		405
Ile His Thr Gly	Glu Lys Pro Tyr Pro	Cys Pro Thr Cys Gly Thr			
	410		415		420
Arg Phe Arg His	Leu Gln Thr Leu Lys	Ser His Val Arg Ile His			
	425		430		435
Thr Gly Glu Lys	Pro Tyr His Cys Asp	Pro Cys Gly Leu His Phe			
	440		445		450
Arg His Lys Ser	Gln Leu Arg Leu His	Leu Arg Gln Lys His Gly			
	455		460		465
Ala Ala Thr Asn	Thr Lys Val His Tyr	His Ile Leu Gly Gly Pro			
	470		475		480

&lt;210&gt; 36

&lt;211&gt; 790

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7510064CD1

&lt;400&gt; 36

Met Ala Leu Gly Leu Gln Arg Ala Arg Pro Ala Leu Ser Cys Gly		
1	5	10
Val Ile Ser Pro Pro Cys Ala Pro Thr Arg Asn Ser His Pro Gly		
	20	25
Pro Gly Cys Thr Ala Ser Pro Pro Ala Pro Pro Gly Trp Pro Phe		
	35	40
Ser Gln Arg Gly Pro Gly Arg Trp Ser Thr Thr Glu Leu Arg Lys		
	50	55
Glu Lys Ser Arg Asp Ala Ala Arg Ser Arg Arg Ser Gln Glu Thr		

	65		70		75
Glu Val Leu Tyr Gln Leu Ala His Thr Leu Pro Phe Ala Arg Gly					
	80		85		90
Val Ser Ala His Leu Asp Lys Ala Ser Ile Met Arg Leu Thr Ile					
	95		100		105
Ser Tyr Leu Arg Met His Arg Leu Cys Ala Ala Gly Glu Trp Asn					
	110		115		120
Gln Val Gly Ala Gly Gly Glu Pro Leu Asp Ala Cys Tyr Leu Lys					
	125		130		135
Ala Leu Glu Gly Phe Val Met Val Leu Thr Ala Glu Gly Asp Met					
	140		145		150
Ala Tyr Leu Ser Glu Asn Val Ser Lys His Leu Gly Leu Ser Gln					
	155		160		165
Leu Glu Leu Ile Gly His Ser Ile Phe Asp Phe Ile His Pro Cys					
	170		175		180
Asp Gln Glu Glu Leu Gln Asp Ala Leu Thr Pro Gln Gln Thr Leu					
	185		190		195
Ser Arg Arg Lys Val Glu Ala Pro Thr Glu Arg Cys Phe Ser Leu					
	200		205		210
Arg Met Lys Ser Thr Leu Thr Ser Arg Gly Arg Thr Leu Asn Leu					
	215		220		225
Lys Ala Ala Thr Trp Lys Val Leu Asn Cys Ser Gly His Met Arg					
	230		235		240
Ala Tyr Lys Pro Pro Ala Gln Thr Ser Pro Ala Gly Ser Pro Asp					
	245		250		255
Ser Glu Pro Pro Leu Gln Cys Leu Val Leu Ile Cys Glu Ala Ile					
	260		265		270
Pro His Pro Gly Ser Leu Glu Pro Pro Leu Gly Arg Gly Ala Phe					
	275		280		285
Leu Ser Arg His Ser Leu Asp Met Lys Phe Thr Tyr Cys Asp Asp					
	290		295		300
Arg Ile Ala Glu Val Ala Gly Tyr Ser Pro Asp Asp Leu Ile Gly					
	305		310		315
Cys Ser Ala Tyr Glu Tyr Ile His Ala Leu Asp Ser Asp Ala Val					
	320		325		330
Ser Lys Ser Ile His Thr Leu Leu Ser Lys Gly Gln Ala Val Thr					
	335		340		345
Gly Gln Tyr Arg Phe Leu Ala Arg Ser Gly Gly Tyr Leu Trp Thr					
	350		355		360
Gln Thr Gln Ala Thr Val Val Ser Gly Gly Arg Gly Pro Gln Ser					
	365		370		375
Glu Ser Ile Val Cys Val His Phe Leu Ile Ser Gln Val Glu Glu					
	380		385		390
Thr Gly Val Val Leu Ser Leu Glu Gln Thr Glu Gln His Ser Arg					
	395		400		405
Arg Pro Ile Gln Arg Gly Ala Pro Ser Gln Lys Asp Thr Pro Asn					
	410		415		420
Pro Gly Asp Ser Leu Asp Thr Pro Gly Pro Arg Ile Leu Ala Phe					
	425		430		435
Leu His Pro Pro Ser Leu Ser Glu Ala Ala Leu Ala Ala Asp Pro					
	440		445		450
Arg Arg Phe Cys Ser Pro Asp Leu Arg Arg Leu Leu Gly Pro Ile					
	455		460		465
Leu Asp Gly Ala Ser Val Ala Ala Thr Pro Ser Thr Pro Leu Ala					
	470		475		480
Thr Arg His Pro Gln Ser Pro Leu Ser Ala Asp Leu Pro Asp Glu					

	485		490		495
Leu Pro Val Gly	Thr Glu Asn Val His Arg	Leu Phe Thr Ser Gly			
	500		505		510
Lys Asp Thr Glu	Ala Val Glu Thr Asp	Leu Asp Ile Ala Gln Met			
	515		520		525
Arg Lys Leu Lys	Leu Arg Leu Leu Thr Thr	Gly Thr Glu Leu Arg			
	530		535		540
Ser Asp Gly Ala	Gly Thr Ser Ala Lys Val	His Pro Ser Pro Arg			
	545		550		555
Leu Ile Leu Leu	Pro Pro Ser Cys Pro	Pro Gln Asp Ala Asp Ala			
	560		565		570
Leu Asp Leu Glu	Met Leu Ala Pro Tyr	Ile Ser Met Asp Asp Asp			
	575		580		585
Phe Gln Leu Asn	Ala Ser Glu Gln Leu	Pro Arg Ala Tyr His Arg			
	590		595		600
Pro Leu Gly Ala	Val Pro Arg Pro Arg	Ala Arg Ser Phe His Gly			
	605		610		615
Leu Ser Pro Pro	Ala Leu Glu Pro Ser	Leu Leu Pro Arg Trp Gly			
	620		625		630
Ser Asp Pro Arg	Leu Ser Cys Ser Ser	Pro Ser Arg Gly Asp Pro			
	635		640		645
Ser Ala Ser Ser	Pro Met Ala Gly Ala	Arg Lys Arg Thr Leu Ala			
	650		655		660
Gln Ser Ser Glu	Asp Glu Asp Glu Gly	Val Glu Leu Leu Gly Val			
	665		670		675
Arg Pro Pro Lys	Arg Ser Pro Ser Pro	Glu His Glu Asn Phe Leu			
	680		685		690
Leu Phe Pro Leu	Ser Leu Ser Phe Leu	Leu Thr Gly Gly Pro Ala			
	695		700		705
Pro Gly Ser Leu	Gln Asp Pro Thr Glu	Leu Thr Gln Phe Leu Leu			
	710		715		720
Ser Val Leu Ser	Phe Pro Ile Leu Asp	Pro Tyr Pro Leu Gly Cys			
	725		730		735
Ala Ala Pro Gly	Leu His Ala Ser Pro	Phe Ser Leu Pro Thr Ile			
	740		745		750
Ser Val Pro Gln	Asn Pro Leu His Phe	Pro Pro Gln Pro Ser Arg			
	755		760		765
His Ala Leu Thr	Leu Thr Leu Pro His	Met Phe Gly Ala Pro Gly			
	770		775		780
Ala Pro Ser Pro	Leu Gly Trp Phe Ala	Ile			
	785		790		

&lt;210&gt; 37

&lt;211&gt; 1154

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7490148CB1

&lt;400&gt; 37

```

agcatcacgt gccagggtgg ggggctataa aatacccgag ccgggcgcgc gcggggggacg 60
tgaggacagc cctctccggg gaccctttg ttcccagccc agacgccaac acctctgcgt 120
ccccaagggc ttgactgccc gtgtctgcgc ggctcccagg gcagagctta gaactactaga 180
ggagaggggt cgccgcgaac tgccggggct tccagccacc caccctctc gacatgtcgc 240

```

```

gtctcttcta tgtcgactcg ctcatcatca aggacacctc acggcctgcg cctcgctgc 300
ctgaaccgca ccccgggcgg gattttcttca tcccgtttgg catgccgccc ccattggtga 360
tgtccgtgtc cgcccccggc tgcccgctccc gcaagagcgg cgcgttctgc gtgtgccctc 420
tctgcgtcac ttcgcacctg cactcctctc ggggggtctgt gggccccgcc agcggggggcg 480
cagggccggg gtttcccggg cccggagaca gtgggggtggc agggccccga ggggcaactgc 540
ctctgcttaa gggccagtgc tcttcggctc ctgggggacgc gcagttttgc ccgcggtga 600
accatgcgca tcatcaccac caccgcccgc agcaccacca tcaccatcat cagccccagc 660
agcctggctc ggccgcgggc gcggcagcag cagcagcggc ggcgggcgcc gcggcggcct 720
tgggggcacc gcagcaccac gcacctgtct gcaccgccac cacctacaac gtggcggaac 780
cgcgagagatt ccactgcctc accatgggag gctctgacgc cagccaggta cccaatggca 840
agaggatgag gacggcgttc actagcacgc aactcctgga gctggagaga gaattctctt 900
ccaacatgta cctgtctcga ctccggagga ttgaaatcgc cacttacctg aacctgtcgg 960
agaagcaggt gaaaatctgg ttccagaacc gccgagtga gcacaagaag gagggggaagg 1020
gcacgcagag gaacagtcac gcgggctgca agtgcgctcg gagccagggt cactacgcgc 1080
gctccgagga tgaggactcc ctgtcgccgg cctcagccaa cgatgacaag gagatttccc 1140
ccttatgagg gagg                                     1154

```

&lt;210&gt; 38

&lt;211&gt; 754

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7490301CB1

&lt;400&gt; 38

```

agccaaaatc tttgaaaggt tctcattgca gatccataat caaaccaccc cagccagaac 60
aattttctgcc tctacagcgc tgtcatggag cagacctgaa tctcacccat ggagacaggc 120
aggcaagcag gtgtgtctgc tgagatgttc gccatgcccc gagatctgaa gggcagcaac 180
aaggatggga tccctgagga cctagatggg aacttggaag aaccagga tcaggaaggt 240
gagctcagga gtgaggatgt catggacctc acagaagggt acaatgaggc ctgagcctca 300
gtcctcctag gagccaaaag acggaaaaca gataccaaag gaaagaagga gaggaagccc 360
actgtggatg cagaggaggc tcagaggatg acaaccctgc tgtctgccat gtctgaggag 420
cagctgtccc gctacgaagt gtgtcgccgg tcagctttcc caaaagcatg cattgcggtg 480
ctgatgcggc ctatcactgg cagatcgggt tctgagaacg tggcgattgc catggctgga 540
atagccaagg tctttgttgg agaggtggtg gaagaggccc tggacgtgtg tgagatgtgg 600
ggagaaatgc cccactgca gcccaagcat ttaaggagg ctgttcgcag gttaaagccc 660
aagggcctct tcccacacag caactacaaa aaaatcatgt tctaggccca aggccagacg 720
gagggtctg tttgtgcagg aataagtacc gcat                                     754

```

&lt;210&gt; 39

&lt;211&gt; 2483

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2383223CB1

&lt;400&gt; 39

```

gtgcgatcgg gttgtgctta gcttgggggc tccctggccc ttgacgcgtc aggttgcgtg 60
acccctgcat cggatgcgct gtacctgcgc ctggctccgt gaaccttagg gacaacaccg 120
ggacaccgca gaggccggaa aatggactca gtggcttttg aggatgtgtc tgtgagcttc 180
agccaggagg agtgggctct gctggctcct tcacagaaga aactctacag agatgtgatg 240
caggaaacat tcaagaacct ggcattctata ggggaaaaat gggaagaccc gaatgttgaa 300

```

```

gatcaacaca aaaaccaagg acgaaatcta agaagccata cgggagagag actctgtgaa 360
ggtaaagaag gtagtcaatg tgcagaaaac ttcagtocca atctcagtgt gacgaagaag 420
actgccggag taaaaccata tgagtgtact atctgtggaa aagccttcat gcgtctctca 480
tccttacta gacacatgag gtctcacact ggatacagagc tatttgagaa gccatataaa 540
tgtaaggagt gtgagaaagc ctttagttat ctcaaatacct ttcaaagaca tgaaaggagt 600
cacactggag aaaaacccta taaatgtaaa caatgtggaa aaaccttcat atatcaccag 660
ccctttcaaa gacatgagcg gactcacatt ggagaaaaaac cctatgaatg taagcaatgt 720
ggaaaagctc ttagttgttc cagttcgctt cgagttcatg aaaggattca cactggagaa 780
aagccctatg aatgtaaaca atgtgggaaa gccttcagtt gttccagttc tattcgagta 840
cacgaaagaa ctcacactgg agagaaaccc tatgcatgta aggaatgtgg gaaagccttc 900
atttcccaca caagtgttct aacacacatg ataacacaca acggagatag acctataaa 960
tgcaaagaat gtggaaaggc attcattttt ccagttttt tacgagtaca tgaaagaatt 1020
cacactggag agaaacccta taaatgtaaa caatgtggta aagccttcag atgttccacc 1080
tccattcaaa ttcataagag aattcatact ggagagaagc cctataaatg taaagaatgt 1140
gggaaatctt tcagtgcacg cccagccttt cgagtacacg tgagagtgc tactggagag 1200
aaaccctata agtgtaaaga atgtgggaaa gcctttagta gaatcagtta ctttcgaata 1260
catgaaagga ctcacactgg agagaaaccc tacgaatgta aaaaatgtgg gaaaactttc 1320
aattatcctc tagatttgaa aatccacaag agaaatcaca ctggagaaaa accctatgag 1380
tgtaaggaat gtgcaaaaac cttcattttt cttgagaact ttcgaagaca catgatcacc 1440
cacactggag acggacctta taaatgtagg gactgtggga aggtgttcat ttttcctagt 1500
gcgttacgaa cacatgaaag aactcacact ggagagaaac cctatgaatg taaacaatgt 1560
ggaaaagcct ttagttgttc tagttacatt cggatacata aaagaactca cactggggag 1620
aaaccttatg aatgtaagga atgcgggaag gcctttatct atcccacaag ctttcaagga 1680
cacatgagaa tgcatactgg agagaaaccc tataaatgta aagaatgtgg gaaggccttt 1740
agtcttcaca gttcctttca aagacataca agaattcaca attatgagaa acctcttgaa 1800
tgtaagcaat gtggaaaagc cttcagtgtg tccacatcct taaaaaaaca tatgagaatg 1860
cacaatcgat agaaactcta taaatgtgag aaataggaga aagttttcaa ttctaacaga 1920
tgctttcaaa gttgtgaaaa ttcccactga agagagaaat cctgtcaatg taagtaatat 1980
agaaagcgag atacaagatg attcatgtat agtcaggtag cacataatca tgtttcagca 2040
gcaatggacc atataggttg gtagcccat aagattatat aacacctaaa acatttctat 2100
caaccgcaat ttagtagcag tagtaacacc atagtgcagc acattattta agtgtttatg 2160
gtgctggtgt aaacgtgctg cactttcagt tgtataaaag tgggaggcca aggcggggcg 2220
cggtggtctc cgcctgtaac cccagcactt tgggaggcca aggcggggcg atcacgaggc 2280
caggagatca agaccatcct ggctaacacg gtaaaacccc atctctacta aaaatacaaa 2340
aaattagccg ggcgtggttg tgggcgcctg tagtcccagc tacttgggag gctgaggcag 2400
gagaatggca taaaccagg aggcgagct gacagtgagc tgagatccgg ccaacagagc 2460
gagactctgt ctcaaaaaaa aaa

```

&lt;210&gt; 40

&lt;211&gt; 2535

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3495982CB1

&lt;400&gt; 40

```

aggggtccaag cgggcccggc cgcgtgggac ggaactggtg ccctccgtgg acaattgtgt 60
tgaagcagaa attgttccgg atctcgggtc ggacacggaa gtcttcctgc agtgtttctg 120
gatgcgggga caggatgag caggaattcc agtctcagtt tccagatgga gcgacccctc 180
gaggagcaag tccagagcaa gtggtcgtct agtcaaggcc gcacaggaac aggaggtct 240
gatgtcctcc agatgcagaa cagtgaacac catggacaaa gcatcaagac tcaaactgac 300
tccatctccc ttgaggatgt ggctgtgaac ttcacctgg aggagtgggc tttgctggat 360
cctggccaga ggaatatcta cagagatgtg atgcgggcaa ccttcaagaa cctggcctgt 420
ataggggaaa aatggaaaga ccaggatatt gaagatgaac acaaaaacca gggaagaaat 480

```

```

ctaagaagtc ctatgggtga agcactctgt gaaaataaag aagattgtcc atgtggaaaa 540
agcactagcc agattcctga tcttaatacg aacctggaaa ctccctactgg attaaaacca 600
tgtgactgca gtgtgtgtgg ggaagtcttc atgcatcagg tctcccttaa taggcacatg 660
agatctcaca ctgaacagaa accaaatgag tgtcacgaat atggagagaa gccacataaa 720
tgcaaagaat gtgggaaaaac cttcactcgc agctccagta ttcgaaccca tgaaagaatt 780
cacactggag agaaacccta tgaatgtaag gaatgtggca aagccttcgc atttctcttt 840
tcctttcgaa accatataag aattcatact ggagagacac cctatgaatg taaggaatgt 900
gggaaggcat tcagatatct cactgctctt cggcgccatg aaaaaaatca cactggagag 960
aaaccctaca aatgtaaaca gtgtggaaaa gcctttatat attaccagcc ttttctaacc 1020
catgaaagga ctcacactgg agagaaacct tatgaatgta agcaatgtgg gaaagccttc 1080
agttgtccca cgtacttacg gagtcatgag aaaactcata ctggagagaa accttttgta 1140
tgtagggaat gtgggagagc cttcttttct cactcaagcc ttogaaaaca cgtgagccac 1200
cacaccgggc cccagttctt tttttttttt tttgagacgg agtccttgcc caggctggag 1260
tgcagtggcg cgtactcagc ttactgcaag ctccgcctcc tgggttcacg ccattctcct 1320
gcctcagcct cccgagtagc tgggactaca ggcgcctcc accacgcccg gctcattttt 1380
tgtattttta gtggagacgg ggtttcacgg tgtagccag gatagtcttg atctcctgac 1440
ctcgtgatcc gccctcctcg gcctcccaa gtgctgggat tataggcgtg agccaccgcg 1500
cccggcctcc agttcatttt aaaaccatga aaccacccac actggagagc cgcttcttta 1560
aatgtaagca atatgggaac atcttcaatg acatctgcaa ccattgtcga gattcttcgt 1620
aggaagaatt aaaggaatgg ggaagaagga cttgggcccc agcaggcctt ttgtggcatc 1680
cagtcagaag ttaggaaaca cctctaggg cagtgcctgc cctcagctct cacacatgac 1740
tgagagagaag tgagaagact cccagaatct gccttttaag tttccatcag tgtcctcggg 1800
tcagactgtg ggaatgctgt acttctaaga caaagggtcg catcttacag tttttcccca 1860
agtgggatct gcctatgcac atcctaattc acagacattg tcagcattgc attattccag 1920
ggctgttttg aggtcatctc cagctccatt tcccatctca aacctacact ttggacactt 1980
gtttaaacat tcacatctga atttctctgt ggcaaccgca ggtggctctg aggtgagaga 2040
ctgaagaagc agagcctgta tcatgggaca cacaccctt gaccacacag catccactct 2100
cctgcggtta ccagccctgc accaaccacc acagtcagaa taatctcaaa actgctagtg 2160
gagtcacaga cacactctgc catgtcttag ctgtaggaca ctcagaatte ttacctcccc 2220
taacttgctt tccctgtttt ctggcacatt ctaccctgag atattcagcg gcatatgcac 2280
tgatgatgct ccggccttct cagtcagtgc catgtcacat gcatcctgct ccaggactgg 2340
tttccatagc tccgctaag actagcatt agctcttctt ggaaaaggac ccaccctgca 2400
gcgctccagc tctgttgtgt atgggaggtg ctggtgtcag ggttcacaag ccatggaact 2460
accttgcttc acatcacaaag gcaagggtgaa ctcaactccc aataaactgc ttgctggagc 2520
ctccgaaaaa aaaaa 2535

```

&lt;210&gt; 41

&lt;211&gt; 2817

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477891CB1

&lt;400&gt; 41

```

gtctttgaaa atacttcatt ttcttagcat ttcaggagat tataacatcc tgtatttcag 60
tttctgagag ctttactgac tgatttccct attcaaaaca atcctcattt cctacatttc 120
tgaagatctc aagatctgga ctactgttga agaaattccc agtaaggctc acttatatct 180
ttagagatgg caaataacta caagaaaatt gttctactga aaggattaga ggtcatcaat 240
gattatcatt ttagaattgt taagtcctta ctgagtaacg atttaaaact taatccaaaa 300
atgaaagaag agtatgacaa aattcagatt gctgacttga tggaggaaaa gttcccagg 360
gatgccgggt tgggcaaact aatagaattc ttcaaagaaa taccaacact gggagacctt 420
gctgaaactc ttaaaagaga aaagttaaaa gtcaaaggaa taatcccatc taaaaagacg 480
aaacagaaaag aagtgtatcc tgctacacct gcatgcaccc caagcaaccg tctcacagct 540
aaaggagcag aggagactct tggacctcag aaaagaaaaa aaccatctga agaagagact 600

```

```

ggaacccaaaa ggagtaagat gtccaaagag cagactcggc cttcctgctc tgcaggagcc 660
agcacgtcca cagccatggg ccgttcccca cctcccaga cctcatcatc agctccacc 720
aacacttcct caactgagag cctaaaacca ttggccaacc gtcacgcaac tgccagtaaa 780
aatattttcc gagaagaccc aataatcgcg atggtactaa atgcaacaaa agtattttaa 840
tatgaatcct cagaaaatga gcaaagaaga atgtttcatg ctacagtggc tacgcagaca 900
cagttctttc atgtgaaggt tttaaacatc aacttgaaga ggaaattcat taaaaagaga 960
atcatcatta tatcaaatta ttccaaacgt aatagtctcc tagaggtgaa tgaagcctct 1020
tctgtatctg aagctgggtc tgaccaaacg tttgagggtc caaaggacat catcagaaga 1080
gcaaagaaaa ttccgaagat caatattctt cacaacaaa cttcaggata tattgtatat 1140
ggattatttta tgctacatac gaaaatttga aataggaaga cgacaatcta tgaaattcag 1200
gataaaacag gaagtatggc tgtagtagga aaaggagaat gccacaatat cccctgtgaa 1260
aaaggagata agcttcgact cttctgcttt cgactgagaa agagggaaaa tatgtcaaaa 1320
ctgatgtcag aaatgcatag tttcatccag atacagaaaa atacaaacca gagaagccat 1380
gactccagga gcatggcact accccaggaa cagagtcagc atccaaaacc ttcagaggcc 1440
agcacaacc tacctgaaag ccactctcaag actcctcaga tgccaccaac aactccatcc 1500
agcagtttct tcaccaagaa aagtgaagac acaatctcca aaatgaatga cttcatgagg 1560
atgcagatac tgaaggaagg gagtcatttt ccaggaccgt tcatgaccag cataggccca 1620
gctgagagcc atccccacac tcctcagatg cctccatcaa caccaagcag cagtttctta 1680
accacgaaaa gtgaagacac aatctccaaa atgaatgact tcatgaggat gcagatactg 1740
aaggaaggga gtcatttttc aggaccgttc atgaccagca taggcccagc tgagagccat 1800
ccccacactc ctcatatgcc tccatcaaca ccaagcagca gtttcttaac caggttgaaa 1860
ccaagactga agactgaacc tgaagaagtt tccatagaag acagtgccca gagtgacctc 1920
aaagaagtga tgggtgctgaa cgcaacagaa tcatgtgtat atgagcccaa agagcagaag 1980
aaaatgtttc atgccacagt ggcaactgag aatgaagtct tccgagtga gggtttttaa 2040
attgacctaa aggagaagtt caccocaaag aagatcattg ccatagcaaa ttatgtttgc 2100
cgcaatgggt tcctggagggt atatcctttc acacttggtg ctgatgtgaa tgctgaccga 2160
aacatggaga tcccaaaagg attgattaga agtgccagcg taactcctaa aatcaatcag 2220
ctttgctcac aaactaaagg aagttttgtg aatggggtgt ttgaggtaaa taagaaaaat 2280
gtaaggggtg aattcactta ttatgaaata caagataata cagggaagat ggaagtgggtg 2340
gtgcatggac gactgaccac aatcaactgt gaggaaggag ataaactgaa actcacctgc 2400
tttgaattgg caccgaaaag tgggaatacc ggggagttga gatctgtaat tcatagtcac 2460
atcaaggtca tcaagaccag gaaaaacaag aaagacatac tcaatcctga ttcaagtatg 2520
gaaacttcac cagacttttt cttctaaaat ctggatgtca ttgacgataa tgtttatgga 2580
tataaggtct aagtgcctaa aaaaatgtac atatacctgg ttgaaatata acactataca 2640
tacacaccac catatatact agctgttaat cctatggaat ggggtattgg gagtgctttt 2700
ttaatttttc atagtttttt ttaataaaaa tggcatagtt ttgcatctac aacttctatt 2760
aattcggaag ataaataact attccagagg ggggggggtc gggggggggg gagagga 2817

```

&lt;210&gt; 42

&lt;211&gt; 2295

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 72688352CB1

&lt;400&gt; 42

```

caggaaagtg caggttcttg gaagtctcgg tgggttcccc gcaaaatcag agatggcatc 60
tcactatgtt gccaggtctg atcttgaact cctgacctca agtgatcctc tctccttggc 120
ctcctaaagt gctgggattc caggtgtgag ccacttcacc cagccacatt tatccatctt 180
gtaaaaatatt gctttgtttt ttggtgcatt tcaaaccac aggtcctgac ttcaggagca 240
agacacttca agtgagggtga ggaggaggtt ccaggacctg gataccatcc tttgcaggga 300
tgtaattcac cagagaccaa gatttctagc cagaaattga gggcagctca aggagagact 360
acagacacag catcctactt actggctatt tcccagtaac aaaagaaaat gatcaagtcc 420
caggaatcac tgaccctgga ggatgtggct gtggagttca cttgggagga gtggcagctc 480

```

```

ctcgccctg ctcagaagga cctgtaccga gacgtgatgt tggagaacta tagcaacctc 540
gtgtcagtgg ggtatcaagc cagcaaacca gatgcactct tcaagttgga acaaggagag 600
ccatggacag tagaaaatga aatccacagc caaatctgtc cagaaatcaa gaaagttgac 660
aatcatctac agatgcactc acaaaagcaa agatgtctga agagagtggg acaatgccat 720
aaacataatg cattttgaaa catcattcat cagaggaaaa gtgattttcc ttttaaggcaa 780
aatcatgata catttgactt acatgggaaa atactgaaat caaattttaag tttagtcaac 840
cagaacaaaa ggtatgaaat caagaattct gtgggggtta atggagatgg gaaatccttc 900
cttcatgcca agcatgaaca atttcataat gaaatgaact tccccgaagg tggaaattct 960
gtgaatacaa attcacaatt cattaagcat cagcgaaactc aaaacataga taaaccccat 1020
gtatgcactg agtgtgggaa ggctttcttc aagaagtctc gcctcatcta tcatcagaga 1080
gttcacactg gggagaaacc tcatggatgc agtatatgtg ggaaagcctt ctccagaaag 1140
tccgggtcct ctgaacacca gagaaaccac acaggagaga aaccctatga atgcactgaa 1200
tgtgacaaag cattccgctg gaaatcacag ctcaatgcac accagaaaat tcatacagga 1260
gagaagtcac atatatgcag tgattgtgga aaaggcttca tcaagaagtc tcggctcatt 1320
aatcatcaga gatttcatac aggagagaaa ccacatggat gcagcctgtg tgggaaggcc 1380
ttctccaaaa ggtccaggct cactgaacac cagagaactc atacaggaga gaaaccctat 1440
gaatgcactg aatgtgacaa agcattccgc tggaaatcac agctcaatgc acatcagaaa 1500
gctcacacag gagagaagtc atatatatgc cgtgattgtg gaaaaggctt cattcagaag 1560
ggaaatctca ttgtacatca gcgaattcat actggagaaa aaccctatat atgcaatgaa 1620
tgtggaaaag gcttcatcca aaagggcaac ctcttattc atcgacgtac tcacactgga 1680
gagaaaccct atgtatgcaa tgaatgtggg aaaggcttca gccagaagac atgtttaata 1740
tcccatcaga gatttcacac aggaaagaca ccctttgtat gtactgagtg tggaaaatcc 1800
tgctcacaca agtcagggtc cattaacac cagagaattc acacaggaga gaaaccctat 1860
acatgcagtg actgtgggaa agctttcaga gataaatcat gtctcaacag acatcggaga 1920
actcatacag gggagagacc gtatggatgc tctgattgtg ggaaagcttt ctcccacttg 1980
tcatgccttg tttatcataa gggaaatgct catgcaagag agaaatgtgt aggttcagtc 2040
aaattgaaaa atccttgctc agagagtcac agcttatcac atacacgtga tctcatacag 2100
gataaagact ctgttaacat ggtgactctg cagatgcctt ctgtggcagc tcagacctca 2160
ttaactaaca gtgcgttcca agcagagagc aaagtagcca ttgtgagcca gcctgttgcc 2220
agaagttcag tctcagcaga tagtagaatt tgcacagaat aaaaaccata tgaatgcaaa 2280
aaaaaaaaa aaaaaa

```

&lt;210&gt; 43

&lt;211&gt; 958

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7490652CB1

&lt;400&gt; 43

```

aatacttgag gagatggata acctattttc catgacatga ttattacgca ttgcatgcct 60
gtatcaaaac atgtagcctt ttggctctgt gagcagcacc atggctgttg gcaagaacaa 120
gcaccttatg aaaggtggca aaaagggagc cgaaaataga gtggttgatc cattttctaa 180
gaaagattgg tgtgatgtaa aagcacttgc tatgttcaat ataagaaata ttggagagac 240
gctagtcacc aggactcgag gaaccaaata tgcactctgat agcctcaagc gtcgtgtgtt 300
tgaagtgaat cttgctgatc tgcagaatga tgaagttgca tttagaaaat tcaagctgat 360
tgctgaagat gttcagaaaa aaactaactt ccaggggcatg gatttaccgg acgaaatgtg 420
ttcgtgggtc aaaaaatggc agaccatgat tgaacctcac attgatgtca agactaccga 480
tggttatattg tttcatctac tctgtgattt tactaaaaaa cacaatctga tacagaaggc 540
ctcttatgct cagcaccaac aggtctgcga aatccagaag aaaatgatgg aaattatgac 600
caaaggtgca aatgacttga aagaagtggc caataaattg attccaggca gcaactggaa 660
agaaaagctt tgccatctca tttatcttct ccatgatgtc tttgttagaa aagtaaaaat 720
gctgaagatg cccaagtttg atttgggaaa attcatgggt aattgttctg gaaaagccac 780
tggggatgag acaggtgcta aagtcgaatt agctgatgga tatgaagcac tgggtccaaga 840

```

atctgttttaa agttcaaaat ttaaaaaaaaaa ggccatcctg ggccataaag caaatctctg 900  
tgaaaaaaat gaaatcatatc aaattatttt ctctgatcat aatggaatta aactaaaa 958

<210> 44

<211> 1978

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7489744CB1

<400> 44

ggaggaacac gctgacacgc gggcggcggt gttataacgg tctaaggtag cgactttatg 60  
aagtcctatc ataaagaaat gaagtaaaat ttcctggaca aatctcccaa tcgctccctt 120  
acatatagtt tgaaatccct tgcccctctc ttccttgtag ttgcttggtg tacacataac 180  
aggttttaaat ccagttctgt acctattcca ggctttcttt ccctttcagc ttactgtggc 240  
tgagaaaaac acataccacc ctaatgacta atctcactcc acaaccattc acaacattga 300  
tccttcaaaa ggtcaggctc tcatactatt ctcccagtta cctaaatggc tattttgtgt 360  
gtcctcttcc ctctcctcct cctcctgtcc ccactcctcct cctgtcccca tcctcctcca 420  
atcttcactt tctgtcatg tctttttatt ttaattcact gagaaaaatg aagtaatagt 480  
agaaaaacta cacctaccta ccagccccc ccccccggtg gcgggagcgg agagcggacc 540  
ccagagagcc ctgagcagcc ccaccgcgc gcggcgcta gttaccatca cccccggga 600  
ggagccgcag ctgccgcagc cggccccagt caccatcacc gcaaccatga gcagcggagg 660  
cgagaccag cagccgccc cggccccccc cggcgcccc gccctcagcg ccgcccagac 720  
caagcccggc actacgggca gcggcgagc gagcggtggt ccgggcgggc tcacatcggc 780  
ggcgcctgcc ggcggggaca agaaggtcat cgcaacgaag gttttgggaa cagtaaaatg 840  
gttcaatgta aggaacggat atgggtttcat caacaggaat gacaccaagg aagatgtatt 900  
tgtacaccag actgccataa agaagaataa ccccaggaag taccttcgca gtgtaggaga 960  
tgagagact gtggagtttg atgttggtga aggagaaaag ggtgcggagg cagcaaagt 1020  
tacaggctct ggtggtgttc cagttcaagg cagtaaatac gcagcagacc gtaaccatta 1080  
tagacgctat ccacgtcgta ggggtcctcc acgcaattac cagcaaaatt accagaatag 1140  
tgagagtggg gaaaagaacg agggatcgga gagtgtctcc gaaggccagg cccaacaacg 1200  
ccggccctac cgcaggcgaa ggttcccacc ttactacatg cggagaccct atgggctgctg 1260  
accacagtat tccaaccctc ctgtgcaggg agaagtgatg gaggggtgctg acaaccaggg 1320  
tgaggagaa caaggtagac cagtgcaggc gaatatgtat cggggatata gaccacgatt 1380  
ccgcaggggc cctcctcgcc aaagacagcc tagagaggac ggcaatgaag aagataaaga 1440  
aaatcaagga gatgagaccc aaggtcagca gccacctcaa cgtcgggtacc gccgcaactt 1500  
caattaccga cgcagacgcc cagaaaaccc taaaccacaa gatgggcaag agacaaaagc 1560  
agccgatcca ccagctgaga attcgtccgc tcccagggtc gagcagggcg gggctgagta 1620  
aatgccggct taccatctct accatcatcc ggtttagtca tccaacaaga agaaatatga 1680  
aattccagca ataagaaatg aacaaaagat tggagctgaa gacctaaagt gcttgctttt 1740  
tgcccgttga ccagataaat agaactatct gcattatcta tgcagcatgg ggtttttatt 1800  
atttttacct aaagacgtct ctttttggtg ataacaaacg tgttttttaa aaaagcctgg 1860  
tttttctcaa tacgccttta aaggttttta aattgtttca tatctgggtc agttgagatt 1920  
ttaagaact tcatttttaa tttgtaataa aagtttaca cttgattttt tcaaaaaa 1978

<210> 45

<211> 2859

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 3363382CB1

&lt;400&gt; 45

```
ccggcgcgatga gcgggcgaggag cgcgtgagcg agccggggcg ggcgaggctgg ggggcccccg 60
agcgcgagcc ggagggcggg ggccgcgagcg cggcgggcggg tggcgggcgag ggcggcgcc 120
agaagatggc ggacggcgac agcggcgagcg agcgcggcggg cggcggtggg cgtgcggg 180
tccagcccg cgtcccgcggc ggcggcgagc aagagacgca ggagctggcc tcgaagcggc 240
tggacatcca gaacaagcgc ttctacttag atgtgaagca gaacgccaaag ggccgcttcc 300
tcaagatcgc cgaggtgggc gcgggcggtt ccaagagccg cctcacgctg tccatggcgg 360
tggccgcccga gttccgcgac tcgctgggcg acttcataga acactacgcg cagctgggccc 420
ctagcagccc cgagcagctg gcgggcgggc cggaggaggg cggcgggccg cggcgcgcg 480
tcaagagcga attcttgggtg cgtgagaacc gcaagtacta cctggacctc aaggagaacc 540
agcgcggccg cttcctgcgc atccgcaaa cgggcaaccg cggcggtggc ggcttcggcg 600
cgggccccgg gccggcggcg ttgcagagcg gccagaccat cgcgctgcct gcgcaggggc 660
tcatcgagtt ccgcgacgcg ctggcgaagc tcatagacga ctacggaggc gaggacgacg 720
agctggcagg cggcccgga ggccggcgccg gggggccagg gggcgccctg tatggagagc 780
tccgggaggg cacctccatc accgtggact ccaagcgctt cttcttcgat gtgggctgca 840
acaaatacgg ggtgtttctg cgagtgcgag aggtgaagcc gtcctaccgc aatgccatca 900
ccgtaccctt caaagcctgg ggcaagttcg gaggcgcctt ttgccggtat gcggatgaga 960
tgaaagaaat ccaggaacga cagagggata agctttatga gcgacgtggg gggggcgagc 1020
gcggcgggcga agagtcagag ggtgaggagg ttgatgagga ttgaaacggg cagcttcccc 1080
tacaggcctc caccacaacca ccacccctt ggctagagaa ttccctttcc tgctcatccc 1140
cagtgagcta gtggagaggg agcaagggag gccgcagagg gaaaaaaca aacgtaacac 1200
agttaagaga aataatcgta agagaacagt gacggacaac ttgagaaaag cagtcaagtt 1260
ccaaggaact gacagcaacc tgcaagagg aaaaacagcat ctcctcacct gcgtaaaatt 1320
gtctcagctt ctgttgtttc tcaactgagg ttctgtaaacc catcaggata atccctggag 1380
ggaatagatc cttgcacatc cagggcaaga aacatgtcca agttaccag accattgata 1440
acagttgcat ttaggttgca cctgggtaat ctggcataaa agatctctct aggcctcact 1500
gttgcggtgt ctatcccttc acctccattg aaatcagcat tttggatcta ggtcttcactg 1560
gaatccttga gaagagaggc ctttacaatt acccagttct gaggggttcag gttcacgaaa 1620
agaaatgcaa cttgggataa tcatgaacag gttaaagata agatttcaag aagccatcta 1680
agaatacaga accaaattgg atccattttt ttaaaaaaat ggttttgcag ggaacctgga 1740
ccaaggcaaa tgtcttttct tcgcagaatt gttttccagg atgccagtgg attcagatag 1800
caatgcttgg agtagaatcc gttactaaaa tagtttcaaa gttgacaaaa aattttcaaa 1860
gataaaagca gttttacatt ggggggttgct gaggtaggca caagaaaaag tcaggcataa 1920
agcacaaggc agactgtttg agtggattgg ttgctgctca cttaaagttgt tccctgatc 1980
tctaaatatg gaggtcatta ccaagaaatg ctttggtatg aatgagagcc agatctccac 2040
tgtgtgagcc agtgaattat ggctaattcg gctgttacag ccactgggtg gctggatttt 2100
aaaccataaa acttgaagat tacctacaaa agtaacagtg tggctataag cctgagcttt 2160
aatggatata catcctcaca gaaaagttgg aaataaccaa aactgaagtc ttaattttacc 2220
ttcagtttaa tctgtggatt tgttcaaaata ctaaagatcc tcagggtccag aattccagca 2280
tcatttattc ttttaaaatt tttagaact tgatccattg tatcagtacc tcacaatcag 2340
agttggcaaa tgatggatga gtgattcaag cagtgcaccc ggtggaagct gaaatccatc 2400
tgtgaatgga actgaagtga acgtgaatat gctgactata tcctggaagc atttttatac 2460
catcttgaaa tttcaacaaa ctggcctttt caggttaatc cagctgtctt tcaagaataa 2520
aagttggggg tttcaaggat cgccctcttct atatttttaa tggattttca gtagaaatga 2580
tttttactaa tcaagttaat cccaccccat caaagggtat tcctagaaat gtcataagacc 2640
taggtaacct tgaattgaat gggagctaac gttctttcca aagttttcag gtattctttg 2700
tgtgacacct tctcaaccag gaggaagta acccgcctc cacaatctta gtattttttt 2760
taaactgcat gcctgcccc tatttgagct gcctttttta tttattgcat atccttttta 2820
ttatcttatt ttggtattat tcaatctata caatctttt 2859
```

&lt;210&gt; 46

&lt;211&gt; 1772

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7491148CB1

&lt;400&gt; 46

```
atgaaggacc acgacgccat caagctcttc gtggggcaga tcccgcgggg cttggacgag 60
caggacctca agccgctgtt cgaggagttc ggccgcacat acgagctgac ggtgctgaag 120
gaccggctca ccggcctcca caaaggctgt gccttcctca cctactgcgc ccgggactct 180
gctctcaagg cccagagtgc actgcacgag cagaagaccc tgccagggtt tcatatcctt 240
aacaacaata acaacaacaa aaaccggcca gaggaccgaa agctgtttgt ggggatgctg 300
ggcaagcagc aggggtgagga ggacgtcaga cgctgttcc agccctttgg ccacatcgag 360
gagtgcacgg tcctgcggag tcctgacggc accagtaaag gctgtgcctt tgtgaagttc 420
gggagtcaag gggaagctca ggcgggccatc cggggctctgc acggcagccg gaccatggcg 480
ggcgctcgt ccagcctcgt ggtcaagctg gcggacaccg accgggagcg cgcgctgcgg 540
cggatgcagc agatggccgg ccacctgggc gccttcacc ccgcgccact gccgctaggg 600
gctgcggcg cctacaccac ggcgatcctg cagcaccagg cgccctgct ggcgggcgga 660
cagggccag gcctaggccc ggtggcgga gtggcgccc agatgcaaca cgtggcgccc 720
tttagcctgg tagctgcgc tctgttgcgc gcggcagcag ccaactcccc gcctggcagc 780
ggccctggca ccctcccagg tcttcggcg cccatcgggg tcaatggagt tcggccctct 840
gacaccccc gatccaatgg ccagccgggc tccgacacgc tctacaataa cgggctctcc 900
ccttatccag cccagagccc cggcgtggct gacccctgc agcaggccta cgctgggatg 960
caccactacg cagcagccta tccgtcggcc tatgccccag tgagcacagc ttttccccag 1020
cagccttcag ccctgcccc gcagcagaga gaaggcccc aaggctgtaa cctcttcac 1080
tatcacctgc ctgaggagtt tggatgatgc gaactcatat agacattcct gccctttgga 1140
gccgttgtct ctgctaaagt ctttgtggat cgagccacca accagagcaa gtgttttggg 1200
tttgtagtt ttgacaatcc aactagtgc cagactgcta ttcaggcgat gaatggcttt 1260
caaattggca tgaagaggct caaggctccag ctaaagcggc ccaaggatgc caaccggcct 1320
tactgacctg ctttactga ccagccacag aaagaaacag aagagtgaga agaaaggaga 1380
ggaaaagcac agaaacgctt gagcagccct tcccgaagga gcagctgcgg acggaggtgg 1440
atcggaacca aggctggtgc ctggggctaa ggccactcta aggattgttt ttatcaagtg 1500
ggtgttctg tgctgcagc atagagcgca ggctggcaga gcaaataagg ctggtgagga 1560
gtgactgtcc aggggaacca gcagagggcg ttgggggtgc caagggcttc tccgcaaggg 1620
aagcccagat ttacttcttt caaaatcata tcatcctta gagtttaggg accaaaggac 1680
tattgctttt ttaagaatat atatatctat ataaattaaa acaaactttc cagcacactg 1740
cgccgttact actgatccgg ctgcctccctc ca 1772
```

&lt;210&gt; 47

&lt;211&gt; 3112

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 8126343CB1

&lt;400&gt; 47

```
atggcgaccg accttcccat catggcgctt ggccccgccc gctccgcgc gcctgcggga 60
gggagcagtt ccgggtgcgg tgcgcgccag gggcgggcg gggcgggcgt cctggccatg 120
gccggcctgt cggacctgga actgcggcgg gagctgcagg ccctgggctt ccagccagga 180
cccatcagg acaccaccg ggatgtctac cgcaacaagc tgccgcgcct cgggggagag 240
gccccgctgc gcgacgagga gcggctgcgg gaggaggccc ggccgcgggg cgaggagcgg 300
ttacgggaag aggcccggtt acgcgaggat gcgcgcctgc gcgccgggc cgccgcggcc 360
tctccgcggg cggagccctg gctctccag ccggcctcgg gctcggccta cgcgaccctt 420
ggggcctacg gtgatatccg gccctccgcg gcttccctgg tagggagccg cggcctcgcc 480
tatcctgccc gccggcgca actcaggcgc cgcgcctcgg tccggggcag ctccgaggag 540
gacgaggacg cccggacgcc cgacagggcc acgcagggcc cgggtctcgc ggcccgccgc 600
tgggtgggag cgtctccgc cccggcgcg ctgccttct ccctcctcgg tcccgaccgg 660
```

cgccccggggcc tgcggggcgac tcgagcggggc cctgctggcg cggcgagggc cgggcctgag 720  
gtggggcgcc ggctggagcg ctggctctct cggcttctgc tctggggccag cctaggggcta 780  
ctgctcgtct tcctgggcat cctttgggtg aagatgggca agccctcagc gccgcaggag 840  
gcggaggaca acatgaagtt attgccagtg gactgtgaga gaaaaacaga tgagttctgt 900  
caggccaagc agaaggcagc cttgctggag ctgctgcatg aactctacaa tttcctggcc 960  
atccaagctg gtaattttga gtgtggaaat ccagagaatc taaaaagcaa atgcattcct 1020  
gttatggaag cccaagaata tatagccaat gtgaccagca gctcctccgc caagtttgaa 1080  
gccgcactga cctggatact gagcagtaac aaggacgtgg gcatctggtt gaaaggagaa 1140  
gaccagtctg aattgggtgac gactgtggac aagggtggtc gcttgggaatc tgcccacccc 1200  
cgcatgggtg ttggctgccg cctgagccgg gccttgctca ctgctgtcac caacgtgtc 1260  
atcttcttct ggtgcttggc ttttttgtgg gggctcctaa ttctcctaaa atatcggtgg 1320  
cgaaagttag aagaggagga acaagccatg tatgagatgg tgaagaagat tatagacgtg 1380  
gtccaggacc attacgtgga ctgggagcag gacatggagc gctatccata tgtaggcact 1440  
ctgcacgtgc gcgacagctt gatccctcca cagagccggg gagtccctgg ctggggcgct 1500  
gggtgtgagg taagatgggg tccgacccca ctgagctccc cagaaggaaa cccagagcct 1560  
tgcatcagat gtcagctgag ggcagtgagc accccagggc caagagtggc tctgctccca 1620  
ctcctgctgc tcccagcat gctcggctct ttatggcagg atgagctaga aaagacatat 1680  
gaaagcacia tcccacacct ctgtagagag aacactggca gtggtgatac cagttaggag 1740  
gccttgggct ccaagtaata gcagctccct ccccagggc acaagtagtg aaggatctcc 1800  
tcagtctctg taagtggag ttcagagctg catgggctcc aggcacagaa cagtcaggtc 1860  
tccggttcca ttgctcttgt gattctctgt gtcccacact cctccatgga tgttatcaac 1920  
agactggggt ccctcatggt tgcagatggt ggccaacagc agctggggca ctgtctgcct 1980  
tgtcacagcc agcaggggaa aagtaggcct attgttgtct tataagcaag agaacatgtt 2040  
cccagtagcc ttcagtgaca cactccttga ggctttattg accagaatct cttacatgtc 2100  
tgtttatgaa tcagtcactg gcaaggagga ggaaattgtt tttcaaactt tttaaacat 2160  
gacctgcagt caaaaacacc tgttacctgc aacatcatgg tccagttccc atacacgcat 2220  
gcgtacactc atgcatatac atgtaattga aagcaaagtt tcacaacatg aatgggggtca 2280  
gcttccaggg cacatagctt catgggaagg ggtggcaggg ccactgtcat gtgcagctgc 2340  
agggcacctt ccacatgggc tgtgactgct ggagttgtac actgggggtcc tgacagaacc 2400  
aggggtggggg agggcagagt cccaggagac agccagcagg gccccccca gtgttgggtc 2460  
tggttgtggt cctccctag cagacacttg ctgtggttcc tggccagatg ctgtcatctg 2520  
ctgggtgatg ctgagagctc acccttgctg agtgcctgct gggagccggg cactacttgg 2580  
agctctccac agatgcaaac tcactcactt ggtcctcacc atggttcacg cacacaacag 2640  
aagtggcccc attttaccta taagggaact gaggcacagg gaagtagagg aatgtgtc 2700  
agagtcccat ggctagtggc atttccgtgc tcccgccacc aggtctact gtcaacatca 2760  
tccctatctg caaccaaga ggaatcagaa aaatagagcc gttagcagtc cactttaagt 2820  
caggcacagt ggctcacacc tgtaatccca gcattttggg aggcocaaagt gggaggattg 2880  
cttgagccca tgagttcaag accagcctgg acaacatagc aaaaccccat ctctacaaaa 2940  
aataaatatt agtgggcatg gtgacacatg cctgtagtcc cagctactca ggaggctgag 3000  
gcaggaggat cactcgagcc caggaggtcg aggcgacagt gagctataac tatgccactg 3060  
cacatcagcc tggacaacag agcaagaccc tatctcttaa aaaaaaaaaa aa 3112

<210> 48

<211> 1939

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7044055CB1

<400> 48

gtggtgaagat ggcggctgtg agtctgcggc tcggcgactt ggtgtggggg aaactcggcc 60  
gatatacctcc ttggccagga aagattgtta atccacaaa ggacttgaag aaacctcgcg 120  
gaaagaaatg cttctttgtg aaattttttg gaacagaaga tcatgcctgg atcaaagtgg 180  
aacagctgaa gccatatcat gtcataaag aggaaatgat aaaaattaac aagggtaaac 240

```

gattccagca agcggtagat gctgtcgaag agttcctcag gagagccaaa gggaaagacc 300
agacgtcatc ccacaattct tctgatgaca agaatcgacg taattccagt gaggagagaa 360
gtaggccaaa ctcagggtgat gagaagcgca aacttagcct gtctgaaggg aaggtgaaga 420
agaacatggg agaaggaaag aagagggtgt cttcaggctc ttcagagaga ggctccaaat 480
cccctctgaa aagagcccaa gagcaaagtc cccggaagcg gggctcgccc ccaaaggatg 540
agaaggatct caccatcccg gagtctagta ccgtgaaggg gatgatggcc ggaccgatgg 600
ccgcgtttta atggcagcca accgcaagcg agcctgttaa agatgcagat cctcatttcc 660
atcatttctt gctaagccaa acagagaagc cagctgtctg ttaccaggca atcacgaaga 720
agttgaaaat atgtgaagag gaaactggct ccacctccat ccaggcagct gacagcacag 780
ccgtgaatgg cagcatcaca cccacagaca aaaagatagg atttttgggc cttgggtctca 840
tgggaaagtg aatcgtctcc aacttgctaa aaatgggtca cacagtgact gtctggaacc 900
gcactgcaga gaaatgtgat ttgttcatcc aggagggggc cgtctggga agaaccgccg 960
ctgaagtcgt ctcaacctgc gacatcactt tcgcctgcgt gtcggatccc aagccggcca 1020
aggacctgtt gctgggcccc agtgggtgtg tgcaagggat ccgccctggg aagtgtctacg 1080
tggacatgtc aacagtggac gctgacaccg tctactgagct ggcccagggt atttgtgtcca 1140
ggggggggcg ctttctggaa gccccgtct cagggaatca gcagctgtct aatgacggga 1200
tgttggtgat cttagcggct ggagacaggg gcttatatga ggactgcagc agctgcttcc 1260
aggcgatggg gaagacctcc ttcttcctag gtgaagtggg caatgcagcc aagatgatgc 1320
tgatcgtgaa catggtccaa gggagcttca tggccactat tgccgagggg ctgaccctgg 1380
cccagggtgac aggccagtc cagcagacac tcttggacat cctcaatcag ggacagtgg 1440
ccagcatctt cctggaccag aagtgccaaa atatcctgca aggaaacttt aagcctgatt 1500
tctacctgaa atacattcag aaggatctcc gcttagccat tgcgctgggt gatgcggtca 1560
accatccgac tcccatggca gctgcagcaa atgaggtgta caaaagagcc aaggcgctgg 1620
accagtctga caacgatatg tccgccgtgt accgagccta catacactaa gctgtcgcga 1680
ccccgccctc accctcccaa tccccctct gacccccctc tcctcacatg gggctcgggg 1740
cctgggagtt cattctggac cagcccacct atctccattt ccttttatac agactttgag 1800
acttgccatc agcacagcac acagcagcac ctttccctg aggccggtgg ggaggggaca 1860
agtgtcagca ggattggcgt gtgggaaagc tcttgagctg ggcactggcc ccccgacga 1920
ggtggctgtg tgttcacac

```

&lt;210&gt; 49

&lt;211&gt; 5902

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7493424CB1

&lt;400&gt; 49

```

gtgggtttgag gctctgcgaa gttacagagg ctccgaggcg cttgcaaaaa tgtggctgtc 60
agagaggggaa gtgcacatga atttgaagac agcattagaa agcgttccgc ccacagtttc 120
tctgtttctca acggctcagt tttacagga taaattttta gtttaagtccc atatgaaggc 180
tcaaaagagc ggtaaagaac aacagcttga cattatgaac aagcagtacc aacaacttga 240
aagtcgtttg gatgagatac tttctagaat tgctaaggaa acggaagaga ttaaggacct 300
tgaagaacag cttactgaag gccagatagc agcaaatgaa gccctgaaga aggattttaga 360
aggtgttctc agtggggttc aagaatacct ggggaccatt aaaggccagg caactcaggc 420
ccagaatgag tgcaggaagc tgcgggatga gaaagagaca ttgttgaga gattgacaga 480
agtcgagcag gagagagacc agctggaaat agttgccatg gatgcagaaa atatgaggaa 540
ggagctttgca gagctagaaa gtgccctcca agagcagcat gaggtgaatg catctttgca 600
gcagaccagc ggagatctca gtgcctatga agctgagcta gaggtctggc taaacctaa 660
ggatgctgaa gcccaaccagc tcaaggaaga gttggaaaaa gtaacaagac ttaccagtt 720
agaacaatca gcccttcaag cagaacttga gaaggaaagg caagccctca agaatgccct 780
tggaagagcc cagttctcag aagaaaagga gcaagagAAC agtgagctcc atgcaaaact 840
taaactcttg caggatgaca ataactgtt aaaacagcaa cttaaagatt tccagaatca 900
ccttaaccat gtgggttgatg gtttgggttcg tccagaagaa gtggcagctc gtgtggatga 960

```

```

gctaagaaga aaactgaaat taggaactgg ggaaatgaac atccatagtc cttcagatgt 1020
cttagggaaa agtcttgctg atttacagaa acaattcagt gaaattcttg cacgctccaa 1080
gtgggaaaga gatgaagcac aagttagaga gagaaaactc caagaagaaa tggctctgca 1140
gcaagagaaa ctggcaactg gacaagaaga gttcaggcag gcctgtgaga gagccctgga 1200
agcaagaatg aatttttgata agaggcaaca tgaagcaaga atccagcaaa tggagaatga 1260
aattcactat ttgcaagaaa atctaaaaag tatggaggaa atccaaggcc ttacagatct 1320
ccaacttcag gaagctgatg aagagaagga gagaattctg gcccaactcc gagagttaga 1380
gaaaaagaag aaacttgaag atgccaaatc tcaggagcaa gttttttggtt tagataaaga 1440
actgaagaaa ctaaagaaaag ccgtggccac ctctgataag ctagccacag ctgagctcac 1500
cattgccaaa gaccagctga agtcccttca tggaaactgtt atgaaaatta accaggagcg 1560
agcagaggag ttgcaggaag cagagagggtt cagcagaaag gcagcacaag cagccagaga 1620
tctcaccgga gcagaagctg agatcgaact cctgcagaat ctctcaggc agaaggggga 1680
gcagtttcga cttgagatgg agaaaacagg tgtaggctact ggagcaaact cacaggctgc 1740
agaaattgag aaactgaatg agacaatgga acgacaaagg acagagattg caaggctgca 1800
gaatgtacta gacctcactg gaagtgcaca caaaggaggc tttgaaaatg ttttagaaga 1860
aattgctgaa ctctgacgtg aagtttctta tcagaatgat tacataagca gcatggcaga 1920
tcctttcaaa agacgaggct attggtactt tatgccacca ccaccatcat caaaagtttc 1980
cagccatagt tcccaggcca ccaaggactc tgggtgtggc cttaagtact cagcctcaac 2040
tcctgttaga aaaccacgcc ctgggcagca ggatgggaag gaaggcagtc aacctcccc 2100
tgccctcagga tactgggttt attctcccat caggagtggg ttacataaac tgtttccaag 2160
tagagatgca gacagtggag gagatagtc ggaagagagt gagctggatg accaagaaga 2220
acccccattt gtgcctctc ctggatacat gatgtatact gtgcttctctg atggttcttc 2280
tgtacccag ggcactggccc tgtatgcacc acctctctcc ttgccaaaca atagccgacc 2340
tctcaccctt ggcactgttg tttatggccc acctctctgt ggggccccca tgggtgtatgg 2400
gcctccacc ccctaacttct ccatcccctt catccctatg ggtgtgtctgc attgcaacgt 2460
ccctgaacac cataacttag agaatgaagt ttctagatta gaagacataa tgcagcattt 2520
aaaatcaaa aagcgggaag aaagggtggat gagagcatcc aagcggcagt cggagaaaaga 2580
aatggaagaa ctgcatcata atattgatga tcttttgcaa gagaagaaaa gcttagagtg 2640
tgaagtagaa gaattacata gaactgtcca gaaacgtcaa cagcaaaagg acttcattga 2700
tggaaatgta gagagtctta tgactgaact agaaatagaa aaatcactca aacatcatga 2760
agatattgta gatgaaattg agtgcatgga gaagactctt ctgaaacgtc gctcagagct 2820
cagggaaagt gaccgactcc tggcagaggc tgagagtgaag agttttattgc aaactgagtc 2940
gacaaaaaat gctgttgaaa agttcactga tgccaaagaga gttaacctcg tcaaagctga 3000
agatgctgag gaattagaaa ggagagctca ggaactgtct gttaacctcg tcaaagctga 3060
tcagcagcta agatcgctcc aggtgatgc aaaggatttg gagcagcaca aaatcaagca 3120
agaagaatc ttgaaagaaa taaacaaaat tgtagcagca aaagactcag acttccaatg 3180
tttaagcaag aagaaggaag aactgacaga agagcttcag aaactacaga aagacataga 3240
gatggcagaa cgcaatgagg atcaccacct gcaggtcctt aaagaatctg aggtgcttct 3300
tcaggccaaa agagccgagc tggaaaagct gaaaagccag gtgacaagtc agcagcagga 3360
gatggctgtc ttggacaggc agttagggca taaaaggag gagctgcatc tactccaagg 3420
aagcatggtc caggcaaaag ctgacctcca ggaagctctg agactgggag agactgaagt 3480
aactgagaag tgcaatcaca ttagggaagt aaaatctctt ctggaagaac tgagttttca 3540
gaaaggagaa ctaaattgtt agattagtga aagaaaaact caacttacac ttataaagca 3600
ggaaattgaa aaagaggaag aaaatcttca ggttgtttta aggcagatgt ctaaacataa 3660
aaccgaacta aagaatattc tggacatgtt gcaacttgaa aacctagagc tacaaggttt 3720
gaagctacaa catgacccaa gggtatctga attagagaag actcaggtgg cagtgtctaga 3780
ggagaaaactg gagttagaga atttgacgca gatatcccag cagcagaaag gggaaataga 3840
gtggcagaag cagctccttg agagggataa acgagaaata gaacgaatga ctctgagtc 3900
ccgagcttta caatcgtgtg ttgagtgttt gagcaaaaga gtttttagcag cagcagaaga 3960
atgtgacatt tgggaaaaaa agttggcaca aacaaaagg gtttttagcag cagcagaaga 4020
aaatagcaaa atggagcaat caaacttaga aaagtggaa ttgaatgtca gaaaactgca 4080
gcaggaacta gaccaactaa acagagacaa gttgtcactg cataacgaca tttagcaat 4140
gcaacagcag ctccaagaaa aacgagaagc agtaaaactc ctgcaggagg aactagctaa 4200
tgtccaagac catttgaaac tagcaaaaca ggacctgctt cacaccacca agcatcagga 4260
tgtgttgctc agtgagcaga cccgactcca gaaggacatc agtgaatggg caaatagggt 4320
tgaagactgt cagaaagaag aggagacaaa acaacaacaa cttcaagtgc ttcagaatga

```

```

gattgaagaa aacaagctca aactagtcca acaagaaatg atgtttcaga gactccagaa 4380
agagagagaa agtgaagaaa gcaaattaga aaccagtaaa gtgacactga aggagcaaca 4440
gcaccagctg gaaaaggaat taacagacca gaaaagcaaa ctggaccaag tgctctcaaa 4500
ggtgctggca gctgaagagc gtgttaggac tctgcaggaa gaggagaggt ggtgtgagag 4560
cctggagaag acactctccc aaactaaacg gcagctttca gaaagggagc agcaattggt 4620
ggagaaatca ggtgagctgt tggccctcca gaaagaggca gattctatga gggcagactt 4680
cagccttctg cggaaccagt tcttgacaga aagaaaagaaa gctgagaagc aggtggccag 4740
cctgaaggaa gcacttaaga tccagcggag ccagctggag aaaaaccttc ttatggcaaa 4800
ccaaaaagat ttggagagaa gacaaatgga aatcagtgat gcaatgagga cacttaaate 4860
tgagggtgaag gatgaaatca gaaccagctt gaagaatctt aatcagtttc ttccagaact 4920
accagcagat ctagaagcta ttttggaag aaacgaaaac ctagaaggag aattggaaag 4980
cttgaaagag aaccttccat ttaccatgaa tgagggacct tttgaagaaa aactgaactt 5040
ttcccaagtt cacataatgg atgaacactg gcgtggagaa gcaactccggg agaaactgcg 5100
tcaccgggaa gaccgactca aggcccaact ccgacactgt atgtccaagc aagcagaagt 5160
attaattaaa ggaaagcggc agacagaggg cactttacac agtttgagga gacaagtaga 5220
tgcttttagg gaattggtca ccagcacctc tgcagattca gcgtcatcac ccagtctgtc 5280
tcagctggag tcttccctca cagaggactc tcaacttggg caaaatcagg aaaagaatgc 5340
ctcagccaga tgaggaatac tgtcttgtgt caaatatatt caaggaaaac actccaatac 5400
tcaactgact cataattggc atgtcccatg gtttttttta atcaagatgc agtgaactga 5460
gaattcggaa actccactgg tagttactgt ggctgtacc atttaatgcc aaggttttat 5520
aaatcactgg tgctagtaca tttgggacca acgtgctatg gggaattaac caacatgggg 5580
tgaagctttt tttaaacctc gtacaggttt taaaacaaag cgggagtttt acgggggtatt 5640
ttaacgtca aaaggggagc cttttttcaa tagaagggtg gttgagggca gcttatatag 5700
aggggagcaa catgaggatg tgggtgacaa acttcttcct gaagaagaag cagaccacaa 5760
actttgggag cgagagaaaa gggagaacaa caatatagga ccagggcgag gtagtatata 5820
ataatgttat catcacacat cgacaaagag aggaataaga tatgatagat ataaaaataa 5880
tgataaatga cgggggggaga aa 5902

```

&lt;210&gt; 50

&lt;211&gt; 2687

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1482140CB1

&lt;400&gt; 50

```

ggagcggcga ctggcgagcc atggcgctgg ggctgcagcg cgcaaggccg gccctttcct 60
gtggagtcac tccaccgccg tgcgcaccca ctcgtaactc gcacccgggt cctggctgca 120
ccgcatcccc tctgcacccc cctggatggc ccttcagcca acggggggcct gggcgatggg 180
cgaccacgga gctgcgcaag gaaaagtccc gggatgcggc ccgcagccgg cgagccagg 240
agaccgaggt gctgtaccag ctggetcaca cgctgccctt cgcccgcggc gtcagcgccc 300
acctggacaa ggctctctac atgcgctca ccatcagcta cctgcgcatg caccgcctct 360
gcgccgcagg ggagtggaac caggtgggag caggggagaa ccactggatg ctgctactga 420
aggccctgga gggcttcgtc atggtgctca ccgcccagg agacatgggt tacctgtcgg 480
agaatgtcag caaacacctg ggctcagtc agctggagct cattggacac agcatctttg 540
atttcaccca cccctgtgac caagaggagc ttcaggacgc cctgacccc cagcagacc 600
tgccaggag gaaggtggag gccccacgg agcgggtgct ctccttgcc atgaagagta 660
cgctcaccag ccgcgggcgc accctcaacc tcaaggcggc cacctggaag gtgctgaact 720
gctctggaca tatgagggcc tacaagccac ctgcgcagac ttctccagct gggagccctg 780
actcagagcc cccgctgcag tgccctggtc tcatctgcga agccatcccc caccaggca 840
gcctggagcc cccactgggc cgaggggcct tctcagccg ccacagcctg gacatgaagt 900
tcacctactg tgacgacagg attgcagaag tggctggcta tagtcccgat gacctgatcg 960
gctgttccgc ctacgagtac atccacgcgc tggactccga cgcggtcagc aagagcatcc 1020
acaccttgct gagcaagggc caggcagtaa cagggcagta tcgcttctcg gcccgagtg 1080

```

```

gtggctacct gtggacccag acccaggcca cagtgggtgc agggggacgg ggcccccagt 1140
cggagagtat cgtctgtgtc cattttttaa tcagccagggt ggaagagacc ggagtggtgc 1200
tgtccctgga gcaaacggag caacactctc gcagacccat tcagcggggc gccccctctc 1260
agaaggacac ccctaaccct ggggacagcc ttgacacccc tggcccccg g atccttgcc 1320
tcctgcaccc gccttccctg agcgaggctg ccctggccgc tgacccccgc cgtttctgca 1380
gccctgacct cgtgcgctc ctgggaccca tcctggatgg ggcttcagta gcagccactc 1440
ccagcacccc gctggccaca cggcaccccc aaagtctctt ttcggctgat ctcccagatg 1500
aactacctgt gggcaccgag aatgtgcaca gactcttcac ctccgggaaa gacactgagg 1560
cagtggagac agatttagat atagctcagg atgctgatgc tctggatttg gagatgctgg 1620
ccccctacat ctccatggat gatgacttcc agctcaacgc cagcgagcag ctaccagg 1680
cctaccacag acctctgggg gctgtccccc gggcccggtg tccggagcttc catggcctgt 1740
cacctccagc ccttgagccc tccctgctac cccgctgggg gaggtagccc cggctgagct 1800
gctccagccc ttccagaggg gacccctcag catcctctcc catggctggg gctcggaaga 1860
ggaccctggc ccagagctca gaggacgagg acgagggagt ggagctgctg ggagtgagac 1920
ctcccaaaaag gtcccccagc ccagaacacg aaaactttct gctctttctc ctccagctga 1980
gtttccttct gacaggagga ccagccccag ggagcctgca ggacccccact gaacttacc 2040
aattccttct ttcatgtctta agttttccca ttctagaccc ctaccctcta ggctgtgctg 2100
ctcctggact tcatgcctct ccattctcat tgcctacaat ctctgtgccc cagaaccccc 2160
tccactcccc accccagccc tccagacatg cacttacctt gactttacc ccatgtttg 2220
gggcacctgg ggctccctca ccccttgggt ggtttgcaat ctgaagactt ctccagccac 2280
acaggcacat gcacaggcac ggtgctgtct gcatattgcc aggtggggag agaagccagg 2340
acccctcagc tgtctgccac catctatgtg cctcccttac cccccagctt tctttctaca 2400
gatggtgcta ctcttggtct cccacaggaa aaggcctccc cccttcttag ccccatttac 2460
cccgtttgtg gaaggcactg ctgcgtctgt tttgtcagag agtggcctat ccagattgg 2520
gctatggggg ggtctgaccc ctccctcctc cctctggagg tgatgtgggc cctcaatgga 2580
gggaattgtg ctgggctagg gaaaggggag ggactagact ggccacactg gctctgaaac 2640
tcaccaacct ctatacacca taaagacctc accttggtag gcaccag 2687

```

&lt;210&gt; 51

&lt;211&gt; 8280

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 394992CB1

&lt;400&gt; 51

```

atggtctccg aggaggagga ggaggaggac ggcgacgccc aggagaccca ggattctgag 60
gacgacgagg aggatgagat ggaagaggac gacgatgact ccgattatcc ggaggagatg 120
gaagacgacg acgacgacgc cagttactgc acggaaagca gcttcaggag ccatagtacc 180
tacagcagca ctccaggtag gcgaaaacca agagtacatc ggcctcgttc tcctatattg 240
gaagaaaaag acatcccgcc ccttgaattt cccaagtcct ctgaggattt aatgggtgcct 300
aatgagcata taatgaatgt cattgccatt tacgaggtag tgcggaactt tggcactgtt 360
ttgagattat ctccctttcg ctttgaggac tttgtgcag ctctggtgag ccaagagcag 420
tgcacactca tggcagagat gcatgttggt cttttgaaag cagttctgct tgaagaagac 480
acttccaata ctacctttgg acctgctgat ctgaaagata gcgttaattc cacactgtat 540
ttcatagatg ggatgacgtg gccagagggt ctgcgggtgt actgtgagag tgataaggag 600
taccatcacg ttcttcttta ccaagaggca gaggactacc catatggacc agtagagaac 660
aagatcaaag ttctacagtt tctagtcgat cagtttctta caacaaatat tgctcgagag 720
gaattgatgt ctgaaggggt gatacagtat gatgaccatt gtagggtttg tcacaaactt 780
ggggatttgc tttgctgtga gacatgttca gcagtatacc atttggaatg tgtgaagcca 840
cctcttgagg aggtgccaga ggacgagtg cagtgtgaag tctgtgtagc acacaagggtg 900
cctggtgtga ctgactgtgt tgctgaaatc caaaaaata aaccatatat tgcacatgaa 960
cctattggat atgatagaag tccgaggaaa tactggttct tgaaccgaag actcataata 1020
gaagaagata cagaaaatga aaatgaaaag aaaatttgggt attacagcac aaaggtccaa 1080

```

```

cttgcagaat taattgactg tctagacaaa gattattggg aagcagaact ctgcaaaatt 1140
ctagaagaaa tgcgtgaaga aatccaccga cacatggaca taactgaaga cctgaccaat 1200
aaggctcggg gcagtaacaa atcctttctg gcggcagcta atgaagaaat tttggaatcc 1260
ataagagcca aaaagggaga cattgataat gttaaaagcc cagaagaaac agaaaaagac 1320
aagaatgaga ctgagaatga ctctaaagat gctgagaaaa acagagaaga atttgaagac 1380
cagtcccttg aaaaagacag tgacgacaaa acaccagatg atgaccctga gcaaggaaaa 1440
tctgaggagc caacagaagt tggggataaa ggtaactctg tgtcagcaaa tcttggcgac 1500
aacacaacaa atgcaacttc agaagagact agtccctctg aaggaggagg ccctgtgggg 1560
tgtctctcag aaacccccga tagcagcaac atggcagaga agaagggtggc atctgagctc 1620
ccccaggatg tgccagtagg tgatttcaaa tcggagaagt ccaacgggga gctaagtga 1680
tctcctggag ctggaaaagg agcatctggc tcaactcgaa tcatcaccag attgcggaat 1740
ccagatagca aacttagtca gctgaagagc cagcagggtg cagccgctgc acatgaagca 1800
aataaattat ttaaggaggg caaagaggta ctggtagtta actctcaagg agaaatttca 1860
cggttgagca caaaaaagga agtgatcatg aaaggaaata tcaacaatta ttttaattg 1920
ggtcaagaag ggaagtatcg cgtctaccac aatcaatact ccaccaattc atttgctttg 1980
aataagcacc agcacagaga agaccatgat aagagaaggc atcttgcaac taagttctgt 2040
ctgactccag caggagagtt caaatggaac ggttctgtcc atgggtccaa agttcttacc 2100
atatctactc tgagactgac tatcaccocaa ttagaaaaca acatcccttc atcctttctt 2160
catcccaact gggcatcaca tagggcaaat tggatcaagg cagttcagat gtgtagcaaa 2220
cccagagaat ttgcattggc tttagccatt ttggagtgtg cagttaaacc agtttgtgat 2280
ctaccaatat ggcgagaatc tttaggacat accaggttac accggatgac atcaattgaa 2340
agagaagaaa aggagaaagt caaaaaaaaaa gagaagaaac aggaagaaga agaaacgatg 2400
cagcaagcga catgggtaaa atacacattt ccagttaagc atcaggtttg gaaacaaaaa 2460
ggtgaagagt acagagtgc aggatattgg ggttggagct ggattagtaa aactcatggt 2520
tataggtttg ttctaaatt gccaggcaat actaatgtga attacagaaa gtcgttagaa 2580
ggaacaaaaa ataatatgga tgaaaatatg gatgagtcag ataaaagaaa atgttcacga 2640
agtcaaaaaa aaataaaaaat agagcctgat tctgaaaaag atgaggtaaa aggttcagat 2700
ctgcaaaaag gagcagacca aaatgaaatg gatattctca agattactga gaagaagac 2760
caagatgtga aggagctctt agattctgac agtgataaac cctgcaagga agaaccaatg 2820
gaagtagacg atgacatgaa aacagagtca catgtaaatt gtcaggagag ttctcaagta 2880
gatgtggtca atgttagtga gggttttcat ctaaggacta gttacaaaaa gaaacaaaaa 2940
tcatccaaac tagatggact tcttgaaagg agaattaaac agtttacact ggaagaaaaa 3000
cagcgactcg aaaaaatcaa gttggagggt ggaattaaag gtataggaaa gacttctaca 3060
aattcttcaa aaaatctctc tgaatcacca gtaataacga aagcaaaaga aggggtgtcag 3120
agtgactcga tgagacaaga acagagccca aatgcaata atgatcaacc tgaggacttg 3180
attcagggat gttcagaaag tgattcctca gttcttagaa tgagtgatcc tagtcatacc 3240
acaaacaaac tttatccaaa agatcgagtg ttagatgatg tctccattcg gagcccagaa 3300
acaaaatgtc cgaaacaaaa ttccattgaa aatgacatag aagaaaaagt ctctgacctt 3360
gccagtagag gccaggaacc cagtaagagt aaaacaaaag gaaatgattt tttcatcgat 3420
gactctaacc tagccagtgc agatgatatt ggtactttga tctgtaagaa caaaaaaccg 3480
ctcatacagg aggaaagtga caccattgtt tcttcttcca agagtgtttt acattcatca 3540
gtgcctaaaa gtaccaatga cagagatgcc acacctctgt caagagcaat ggactttgaa 3600
ggaaaactgg gatgtgactc tgaatctaata agcactttgg aaaatagttc tgataccgtg 3660
tctattcagg atagcagtga agaagatatg attgttcaga atagcaatga aagcatttct 3720
gaacagttca gaactcgaga acaagatgtt gaagtcttgg agccgttaaa gtgtgagttg 3780
gtttctgggt agtccactgg aaactgtgag gacaggctgc cgggtcaagg gactgaagca 3840
aatggtaaaa aaccaagtca gcagaagaaa ttagaggaga gaccagttaa taaatgtagt 3900
gatcaaataa agctaaaaaa taccactgac aaaaagaata atgaaaatcg agagtctgaa 3960
aagaaaggac agagaacaag tacatttcaa ataaatggaa aagataataa acccaaaata 4020
tatttgaaag gtgaatgctt gaaagaaatt tctgagagta gagtagtaag tggtaatgtt 4080
gaaccaaagg ttaataatat aaataaaata atccctgaga atgatattaa atcattgact 4140
gttaagaat ctgctataag gccattcatt aatgggtgat tcatcatgga agattttaat 4200
gaaagaaaca gctccgaaac aaaatcgcat ttgctgagtt cttcagatgc tgaaggtaac 4260
taccgagata gccttgagac cctgccatca accaaagagt ctgacagtac acagacgacc 4320
acaccctcag catcttgctc agaaagcaat tcagttaatc aggtagaaga tatggaaata 4380
gaaacctcag aagttaagaa agttacttca tcacctatta cttctgaaga ggaatctaata 4440

```

```

ctcagtaatg actttattga tgaaaatggg ctgccccatca acaaaaatga aaatgtcaat 4500
ggagaatcta aaagaaaaac cgtcatcaca gaagtcacca cgatgacctc cacagtggcc 4560
acagaatcaa aaactgtgat caaggtagaa aaaggcgata agcaaactgt ggtttcttcc 4620
acagaaaatt gtgcaaaatc cactgtcaca accaccacta caacagtgc caagctttcc 4680
acaccctcca caggcggcag tgtggacatc atctctgtaa aggagcagag caaaaccgtg 4740
gtcaccacga cagtgcacaga ctccctgacc accacgggag gcacactggg tacatctatg 4800
actgtgagca aagagtattc cacacgagac aaagtgaac tgatgaaatt ttcaagacca 4860
aagaagactc gttcagggtac agctctgcca tcctatagaa aatttggtac caagagcagc 4920
aagaagagca tttttgtttt gcctaattgat gacttaaaaa agttggcccg aaaaggagga 4980
atccgagagg tcccttattt taattacaat gcaaaacctg ctttgatat atggccatat 5040
ccttctccta gaccgacctt tggcatcact tggagggtata gacttcagac agtaaaagtc 5100
ttagtggag tgagcctgat gttacgggta ctgtgggcaa gtttgagatg ggatgatatg 5160
gcgccaagg ctctccagg aggagggact acacgggacag aaacatccga aactgaaatc 5220
acaacaacag aaataattaa gaggagagat gttggctcct atggcattcg atctgaatat 5280
tgtatcagga aaatcatttg tcccattgga gttccagaaa caccaaaaga aacgcctaca 5340
cctcagagga aaggccttcg atcaagtgc ctgcgggcaa agagaccaga aacgccaag 5400
caaactggcc ctgttattat tgaaacctgg gtagcagaag aagaactgga attgtgggag 5460
atcagggcat ttgctgagag agtggagaaa gaaaaggcac aagcagttga gcaacaggct 5520
aagaaacgac tggagcagca gaagccgaca gtgattgcaa cttccactac tccccaca 5580
agcagtacaa ccagcaccat ctctccagca cagaaggtta tgggtggcccc cataagtggc 5640
tcagttacaa ctggaaccaa aatggtacta actactaaag ttggatctcc agctacagta 5700
acattccaac aaaacaagaa ctttcatcaa acctttgcta catgggttaa gcaaggccag 5760
tcaaattcag gcgttggtca agtacagcag aaagtcctgg gtatcattcc atcaagtaca 5820
ggtaccagtc agcaaacctt tacttcattc cagcccagga cagcaacagt cacaattagg 5880
cccaatacct caggctctgg aggaaccaca agcaattcac aagtaatcac agggcctcag 5940
attcgccctg gtatgaccgt gattagaaca ccactccaac agtcaacact aggaaaggca 6000
attattcgaa cacctgtgat ggtacagcca ggtgctcctc agcaagtgat gactcaaat 6060
atcagggggc agcctgtctc cactgcagtc tccgccccta acacggtttc tcacaacct 6120
gggcagaaaa gcttaacttc agcaacgtcc acttcaata tacagtcttc agcctcaca 6180
ccccctcgcc cccaacaagg acaagtgaag ctaccatgg ctcaacttac tcagttaaca 6240
cagggccacg gtggcaatca aggtttgaca gtagtaattc aaggacaagg tcaaactact 6300
ggacagttgc agttgatacc tcaaggggtg actgtactcc caggcccagg ccagcagcta 6360
atgcaagctg caatgccaaa tgggtactgt cagcgattcc tctttacccc attggcaaca 6420
acagccacca cagccagcac caccaccacc actgtttcca cgacagcagc aggtacaggt 6480
gaacaaaggc agagtaaact gtcaccccag atgcaggtac atcaagacaa aaccctgcc 6540
ccagctcagt catcaagtgt ggggtccagca gaagcccagc cacagactgc tcagccttca 6600
gtcagcccc agcccaaac ccagcccag tcccagctc agcctgaagt tcagactcag 6660
cctgaagttc agacccaaac aactgtttca tcccagctc cttctgaagc acaaccacc 6720
cacgcacagt catccaagcc ccaagttgca gcacagtctc agcctcaaag taatgtccaa 6780
ggacagtctc ctgttcgtgt ccaaagteca tcacagactc gaatacgtcc atcaactcca 6840
tcccaactgt ctctggaca acaatcccag gttcagacta caacctcaca accgattcca 6900
attcaaccac atacatctct tcagatacct tcccaaggcc agccacagtc acaaccacc 6960
gtggtgatga agcataatgc tgtaatagaa catttaaaac agaaaaagag catgactcca 7020
gctgaaagag aagagaatca aagaatgatt gtctgtaacc aggtgatgaa gtatatattg 7080
gataagatag ataaagaaga aaaacaggca gcaaaaaaac ggaagcgtga agagagtgtg 7140
gagcagaaac gtagcaagca gaatgccact aactgtcag ctctgctctt caagcacaaa 7200
gagcagctca gagccgagat cctgaagaag agagcactcc tggacaagga tctgcaaatt 7260
gaagtgcagg aagagctgaa gagagacctg aaaattaaga aagaaaaaga cctgatgcag 7320
ttggctcagg ccacagcagt agctgcaccc tgccccccag tgacaccagc tctccagcc 7380
cctccagccc ctccacctc acctccccct ccacctgctg tgcaacacac aggccttctg 7440
tccacgccc cttacctgc tgcctcccag aagaggaagc gggagagga aaaagactcc 7500
agctcaaagt ccaagaaaaa gaaaatgatc tctactacct caaaggaaac taagaaggac 7560
acaaagcttt actgtatctg taaaacgcct tatgatgaat ctaagttcta tattggctgt 7620
gatctttgta ctaactggta tcatggagaa tgtgttggca tcacagaaaa ggaggctaag 7680
aaaatggatg tgtacatctg taatgattgt aaacgggcac aagagggcag cagtaggaa 7740
ttgtactgta tctgcagaac accttatgat gagtcacaat tttatattgg ctgtgatcgg 7800

```

tgtcagaatt ggtaccatgg gcgctgcgtt ggcattcttc aaagtgaggg agagctcatt 7860  
gatgagtatg tctgtccaca gtgccagtca acagaggatg ccatgacagt gctcacgcca 7920  
ctaacagaga aggattatga ggggttgaag aggggtgctcc gttccttaca ggcccataag 7980  
atggcctggc ctttccttga accagtagac cctaattgatg caccagatta ttatggtgtt 8040  
attaaggaac ctatggacct tgccaccatg gaagaaagag taaaaagacg atattatgaa 8100  
aagctgacgg aatttgtggc agatatgacc aaaatctttg ataactgtcg ttactacaat 8160  
ccaagtgact ccccatctta ccagtgtgca gaagttctcg aatcattctt tgtacagaaa 8220  
ttgaaaggct tcaaagctag caggtctcat aacaacaaac tgcagtctac agcttcttaa 8280

<210> 52

<211> 2597

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 5093550CB1

<400> 52

cctgcaagca ttgggagatc cacagctaag acgccagggc ctcccctgga agcctagaaa 60  
tgggaccatt gacatttatg gatgtggcca tagaattctc tgtggaggag tggcaatgcc 120  
tggaactgac acagcagaat ttatatagga acgtgatgtt agagaactac agaaacctgg 180  
tcttctctggg tattgtctgc tctaagccag acctgatcac ctgtctggag caaggcaaag 240  
agccctggaa tatggagaga catgagatgg tggccaaacc cccaggtatg tgttggtatt 300  
ttgcccaaga ccttcggcca gagcagagca taaaagcttc tttgcaaaga ataatactga 360  
gaaaatatga aaaatgtgga catcacaatt tacagttaaa aaaaggctat aaaagtgtgg 420  
atgagtacaa ggtgcacaaa ggaagtata atggatttaa ccagtgtttg acaactaccc 480  
agagcaaaat atttcagtgt gataaatatg tgaaagactt tcataaattt tcaaattcaa 540  
atagacataa gactgaaaag aatcctttca aatgtaaaga atgtggcaag tcattttgtg 600  
ttctttcaca cctaactcaa cataaaagaa ttcacactac tgtcaattcc taaaacttg 660  
aagaatgtgg caaggccttt aatgtgtcct caacccttc tcaacataag agaattcata 720  
caggacaaaa aactacaaa tgtgaagaat gtggcatagc cttaacaag tcctcacacc 780  
ttaacacaca taagataatt catactggag agaaatccta caaacgtgaa gaatgtggaa 840  
aagcttttaa catatcctca caccttacta cacataagat aattcatact ggagagaatg 900  
cctacaaaatg taaagaatgt ggcaaagctt ttaaccaatc atcaaccctt actagacata 960  
agataaattca tgctggagag aagccctaca tatgtgaaca ttgtggcaga gcttttaacc 1020  
aatcctcgaa cttactaaa cataagagaa ttcatactgg tgataaacct tataaatgtg 1080  
aagaatgtgg caaagccttt aatgtgtcct caacccttac tcaacataag agaattcata 1140  
ctggagagaa acctacaaa tgtgaagagt gtggcaaagc cttaacgtg tcctcaactc 1200  
ttactcaaca taagagaatt catactggag aaaaaccata caaatgtgaa gaatgtggca 1260  
aagcctttta cacatcctca cacctcacca cacataaag aattcatacc ggagagaaac 1320  
cctacaaatg tgaagaatgt ggcaaagcct ttaaccagtt ctcaacaact actacacata 1380  
agataattca tactggagag aaaccctaca aatgtaaaga atgtggcaaa gcttttaagc 1440  
ggtcctcaaa cttactgaa cataggataa ttcatactgg agagaaacc taaaatgtg 1500  
aagaatgtgg caaagctttt aacctatcct cacaccttac aacacataag aaaattcata 1560  
ctggagagaa accctacaaa tgtaaagaat gtggcaaagc ttttaaccaa tcctcgacac 1620  
ttgctagaca taagataatt catgctggag agaaacccta caaatgtgaa gaatgtggca 1680  
aagcttttta ccaatactca aaccttactc aacataagat aattcatact ggagagaaac 1740  
cctacaaatg tgaagaatgt ggcaaagcct ttaattgggtc ctcaactctg actaaacata 1800  
aatgctcaaa cttactaca cacaagaaaa ttcatgtgtg agaaaaatct gacaaataag 1920  
aaaaatggac caggcgtggt ggctcacgcc tgtaatccca gcactttggg aggccgaggc 1980  
gggtgggatca caagatcaag agattgagac catcctggcc aacgaagaga attactggag 2040  
agaaaccgta caaacctgaa agatataaca gtgcagttga caacacctca atatttctaa 2100  
acacctttta atggttgtca tacttgattg taggtaaaat atagtagagg aaacctctac 2160  
aagtgtgaag aatgtggcaa aacttttaac caatgtgtac actttattgg acaggaaatg 2220

```

atgtatactt gagaaaaatt atgcaaagt aaagaatgtg aagaagccat taatatctgt 2280
tcacatctta ttcaatatca gagagttcat acttaataga agcattaaag atgaaattac 2340
tgtcaataga tttttcagaa aatataagcc tttaaagtga agaagagtat ttattttgaa 2400
gacaagcatt aaaatatcaa gagtgtaagc tgggcatggg ggctcaggcc tgtaatttca 2460
gcacttggag gccggggcag gtggtttact tgaggtcaga gggttcagacc agcctggcca 2520
acatggtgaa actccatctc taccaggat gcagaaatta ggtgggtgtg gtgggtgcatg 2580
actaatccca gctactc                                     2597

```

&lt;210&gt; 53

&lt;211&gt; 725

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7487977CB1

&lt;400&gt; 53

```

gcatgtccaa gccggtggac cacgtcaagc ggcccatgaa cgccttcatg gtgtgggtcgc 60
gggctcagcg gcgcaagatg gcccaggaaa accccaagat gcacaactcg gagatcagca 120
aacgcctagg tgccgaatgg aagcttctgt ccgaggcaga gaagcggcca tacatcgatg 180
aagccaagcg gctacgcgcc cagcacatga aggagcacc tgactacaag taccgacctc 240
ggcgcaagcc caagaacctg ctcaagaagg acaggtatgt cttccccttg ccttacctgg 300
gcgacacgga cccgctcaag gcggtgggccc tgcccgtggg ggctctcgac ggctctctga 360
gcgcgcccga gaaagcccgg gccttcttgc cgccggcctc ggcgccctac tccctgctgg 420
accccgcgca gtttagctcg agcgccatcc agaagatggg cgaagtggcc cacaccttgg 480
ctaccggcgc tctgccctac gcgtccaccc tgggctacca gaacggcgcc ttccggcagcc 540
tcagctggcc cagccagcac acgcacacgc acccgtccc caccaacct ggctacgtgg 600
tgccctgtaa ctgtaccgcc tgggtctgct ccacctgca gcccccgtc gcctacatcc 660
tcttcccagg catgaccaag actggcatag acccttattc gtcagcccac gctacggcca 720
tgtaa                                     725

```

&lt;210&gt; 54

&lt;211&gt; 1952

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1706514CB1

&lt;400&gt; 54

```

gcggcccgagg agtacctgta cttttcagct gcgcgggccc cgaggccacg gagagctcgc 60
cttggagagc ccaggagcag gggagacatg gactcagtg tctttgagga tgtggctgtg 120
gacttcaccc tggaggagtg ggctttgctg gattctgctc agagggacct ctacagagat 180
gtgatgctgg agaccttccg gaacctggcc tcagtagatg atgggactca atttaaagcc 240
aatgggtcag tttctctgca ggatatgtac gggcaagaaa aatctaagga acagacaata 300
ccaaacttca caggaaataa ttcctgtgcc tacactttag aaaaaaattg tgaaggctat 360
ggcactgaag accaccacaa aaatctgaga aatcatatgg tggacagatt ctgtacacat 420
aatgaaggta atcaatatgg agaagccatc catcaaatgc cagatcttac cctgcacaag 480
aaagtctctg ctggagaaaa accatatgaa tgcaccaagt gcaggacagt cttcacgcat 540
ctttcttctc ttaaaaggca cgtcaagtct cactgtggac gaâaagcacc tccaggtgag 600
gaatgtaagc aggcctgcat ttgtccctca cacctacaca gtcacggaag aaccgacact 660
gaggagaagc cgtataagtg tcaagcatgt gggcaactt tccaacatcc tcgttacctc 720
tcccaccacg taaagactca cacagcagag aaacctaca aatgcgagca gtgtcggatg 780
gcgtttaatg gggtcgcaag cttcactaga catgtgagaa ctcacacaaa agacaggcca 840

```

```

tatpaatgtc aggaatgtgg gagagccttc atttatccct cgacatttca aagacacatg 900
acaacacaca ctggagagaa gccctataaa tgtcagcact gtgggaaagc cttcacttac 960
ccccaggctt ttcaaagaca tgagaagacg cacacgggag agaagcccta tgaatgcaag 1020
cagtgtggga aaacattcag ttggtctgaa accttgcgag tccacatgag gatccacacc 1080
ggggacaaac tctataaatg tgaacactgt gggaaggctt ttacctcttc cagatcattc 1140
caaggtcatt tgaggacgca cactggagag aaaccttatg agtgtaaaca atgtgggaaa 1200
gccttcactt ggctctcaac gtttagagaa catgtgagaa ttcacacgca agagcagctc 1260
tataaatgtg aacaatgtgg gaaggctttt acctcttcca gatcattccg aggtcatttg 1320
aggacgcaca ctggagagaa gccttatgag tgtaaacaat gtggaaaaac cttcacttgg 1380
tcctcaacgt ttagagaaca tgtgagaatt cacacgcaag agcagctcca taaatgtgaa 1440
cactgtggga aggcctttac ctcttcaga gcattccaag gtcatttgag gatgcacact 1500
ggagagaagc cttatgagtg taaacaatgt ggaaaaacct tcacttgggtc ctcaacctta 1560
cataatcatg tgaggatgca cactggagag aaacctcaca aatgtaaaca atgtgggatg 1620
tccttcaagt ggcactcctc cttccggaac catctgagga tgcacacagg acagaaatcc 1680
cacgaatgtc agtcatactc aaaagccttc agttgccaaag tcattctttc taaaaccagt 1740
gagagcacac actaaagaga aattctataa ctttaatggg gtaacctcac attaatcat 1800
gtataatgct ccagaaaatt cacaccagga gagaaatctt acaagtatga tattgtcttt 1860
gtcaatacct catgtgtaaa acagacccat tagtcgtgaa tttccagctc tttcaaaaat 1920
aaatgtttat gtgagaaaaa aaaaaaaaaa gg 1952

```

&lt;210&gt; 55

&lt;211&gt; 1213

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; In:byte ID No: 7488247CB1

&lt;400&gt; 55

```

gttttgcctc gcccgccga gctgcaggga ccgcacgggtg cccgggtctc ccgcccgcaga 60
ggatccgctg gcccgggccg ggggagcccg cgctgaccac cgcgcgctgt ccccgaggc 120
agccggcctg ccgccatggc cgaccacctg atgctcgccg agggctaccg cctgggtgcag 180
aggccgccgt ccgcgcgggc cgcccatggc cctcatgggc tccggaactc gccgcgtac 240
gcgggcccg gcctggacag tgggctgagg ccgcgggggg ctccgctggg gccgcgccg 300
ccccgccaac ccggggccct ggcgtaacgg gccttcggggc cgccgtctc cttccagccc 360
tttccggccg tgctccgccc ggccgcgggc atcgcgccac tgcagcctgt ggcgacgccg 420
taccgccggc gcgcggccgc gccccccaac gtcggggag gccccccggg cccgcagccg 480
gcgccaagcg ccgcagcccc gccgcgccgc gcgcacgccc tgggcggcat ggacgcccga 540
ctcatcgacg aggaggcgct gacgtcgctg gagctggagc tcgggctgca ccgcgtgccc 600
gagctgcccg agctcttctt gggccagagc gagttcgact gcttctcgga cttgggggtcc 660
gcgcgcgccg ccggctccgt gagctgctga gggcgccggc cgcccgcccg gcgtgccgga 720
gaggagaaag ggcccgaact cccgcgggac cctgcaccca gcgactgggc cccgcgcgcg 780
ccctccgcga ggggtggaggc ggcggtctgt tgccgagggc ccggcaccgg actgggacct 840
tgggctccct ccaggccttg cctcctgcgg gaggacagtt tggcttact tctctgacct 900
cagcctcggc cgtaaagtga aagagaccgg accagcttca gctttcggac tctgttctt 960
ggatcggtgc ctctccccct cgccgccctc ttcccccaat ctgagccatt gcaggcctct 1020
gcctgctgcc cctctctctc tcgggatcgg gtccccagag ccaccatctc ctgagcctcc 1080
caccgccgtg cctgggccct gtggttgctg ggcctccac ctcaaggagg ggaagggttg 1140
acagcccga cccgtggagc aatgcctgt ctggcctcca aaacaaaat aaaactgggt 1200
cactttaaaa aaa 1213

```

&lt;210&gt; 56

&lt;211&gt; 2257

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 1427269CB1

<220>  
<221> unsure  
<222> 2238  
<223> a, t, c, g, or other

<400> 56  
gcggggcggt tatggcggt ccatattaac agcctcctcc tcctccgccc cgcgcgcgt 60  
ctcctcctcc tcctcctttc cctcccgcgc gcgctctaag ccatctccgc cttcaccctg 120  
acgcctgcct cttcccctca cctttccccc tcccctgttc taccatgccc ggcatgatgg 180  
agaaagggcc cgagttactg gggaagaacc gatcgcccaa cggcagcgcc aagagcccgg 240  
caggcggcgg cggcagcgcc gcctcgtcca ccaacggcgg gctgcactac tcagagcccc 300  
agagcggtcg cagcagcgac gacgagcagc atgttgggat gagagtcgga gccgaatacc 360  
aagctcggat ccctgaattt gatccagggt ctacaaagta cacagataaa gacaatggag 420  
ggatgcttgt atggtctcca tatcacagta tcccagatgc caaattggat gaatacattg 480  
caattgcaaa ggaaaagcat ggctacaatg tggaacaggc acttggcatg ttgttctggc 540  
ataaacataa cattgagaag tcccttgctg atctccctaa tttcactccc tttccggatg 600  
agtggacagt ggaagataaa gtcctatttg aacaagcctt tagttttcat ggaaagagct 660  
ttcacaggat tcagcaaag cttccagata agacaattgc aagccttgta aaatattact 720  
attcttgga aaaaactcgc tctaggacaa gtttgatgga tcgccagggt cgtaaactag 780  
ctaatagaca taatcagggt gacagtgatg atgatgtaga agaaacacat ccaatggatg 840  
ggaatgatag tgattatgat cccaaaaaag aagccaaaaa agagggtaat actgaacaac 900  
ctgtccaaac tagcaagatt ggacttgga gaagagagta tcagagtta caacatcgcc 960  
atcattctca gcgttctaag tgccgtccac ctaagggtat gtatttaacc caggaagatg 1020  
tggtagcagt ttcctgtagt ccaatgcag ccaacaccat cclgaggcaa ctggacatgg 1080  
agttgatctc tctaaaacgt caggttcaga atgctaagca agtaaacagt gcacttaaat 1140  
agaaaatgga aggtggaatt gaagaattca aacctcctga gtcaaatacag aaaattaatg 1200  
cccgttgga cagagaggag cagcttctag cagtgcagg tgccgcaaa tatggtaaag 1260  
atttcaagc tattgcagat gtaattggca acaagactgt tggccaagt aagaacttct 1320  
ttgtaacta caggcgctcg ttttaacttag aggaggtatt gcaggagtgg gaagcagaac 1380  
aaggaaccca ggcttcta at ggtgatgctt ctactttagg ggaggagaca aaaagtgtct 1440  
ctaattgtcc atcagggaag agcactgatg aagaagagga ggcacagacc ccacaggctc 1500  
ctcggacact ggggtccatca cctcctgccc catcatccac tccaacacca acagccccta 1560  
ttgccactct gaaccagcct ccaccacttc ttctgccaac actgcctgct gcccgggctc 1620  
ttcaccggca gcctcctcca ctccagcagc aggtcgggtt catccagccc cggccaactt 1680  
taaatacagc tccaccacct cttattcgcc ctgctaattc catgccacc cgtctaaacc 1740  
caagaccggt gttgtccacg gttggtggtc aacagccacc atcacttatt ggaattcaga 1800  
cagattcaca gtccctactg cactaaaaat taaattggac acagctgcag taacttttca 1860  
ccccatcatt ataccagtgc tcatctgact gatgaaaaag aggaaagaat aatcatttct 1920  
agatactgag gctgcgaact agttctgtgg cagtggacta gcataagtgg atgtctaaga 1980  
aattttttag ttacttagac taaaatgttt tacaacaaaa agcctccagt tagcctcctt 2040  
tctagagtat atgttcagca atgtgatctc ataaaaggaa aaaccaaaga tttagtgttc 2100  
tatataccaa ggttttggtt tgttttactg gatttattta atggaggtct ttattatcct 2160  
ggctctcata gtcaaggctc tagtacagga tattgctttt aggatgtgaa acctcttaag 2220  
gtcttaaggt aggatgtngg ctttttcttt aattttt 2257

<210> 57  
<211> 2415  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature

&lt;223&gt; Incyte ID No: 103135CB1

&lt;400&gt; 57

```
gcgggcgcc tggacggcct ggaaggccag cgcgcaccac cgagacgtgg gctcctagag 60
gggcccgaag ctttcggata acaacttcog ctccgggaagt ttgtaaaagt ctgggctacc 120
ggcgcgccgt agtggatgca gcatcctagt ggaggacgcc cctgtgatct gccctccttg 180
gcaactgtgt tccccagagg ggtggcctcg ctgttcccat ggacatggcc caggagccag 240
tgaccttcag ggacgtggcc atctacttct caagggagga gtgggctgtg ctggaacca 300
gccagagggc cctctaccgg gacgtgatgc tggacaactt cagcagtgtg gctgctctgg 360
gattttgcag ccccagacca gacctcgtct ctccgctgga acagtgggag gagccgtggg 420
ttgaagaccg ggagagacct gagttccagg cagtgcagag gggaccccg ccaggggcaa 480
ggaagtctgc agaccccaag agacattgtg atcatccagc ttgggctcac aagaaaacc 540
acgtgcggcg agaaagagcc aggggaaggaa gcagctttag gaaggcttc aggtctggaca 600
cggatgacgg gcagcttccc agagctgtct cagaaaggac agacgccaag cccacggctt 660
tcccgtgtca ggtgctcacg cagcgttgtg ggcggcgcc gggcgcgaga gagcgccgga 720
agcagcgcg agtagagctg tcgttcactc gcggcacgtg cgggaaggcg ctgagctgcc 780
acagccggct gctcgctcac cagacggtgc acacgggaac caaggccttc gagtgcctcg 840
agtgcggcca gaccttcggg tgggcttcaa acctgcagcg ccaccagaag aaccacacgc 900
gcgagaagcc cttctgctgc gaggcctgcg ggcaggcggt cagcctgaag gaccgcctgg 960
ctcagcaccg caaggtccac accgagcaca ggccctactc gtgtggcgac tgtgggaaag 1020
ccttcaagca gaagtccaac cttctcagac accagctggt gcacaccggg gagcgcccg 1080
tctactgctg ggactgcggc aaagccttcc ggaccaagga gaacctcagc caccaccaga 1140
gggtccacag cggggagaag ccctacacct gtgccgagtg cggcaagtcc tcccggtggc 1200
ccaagggctt cagcatccac cggaggctgc acctgacgaa gaggttctac gagtgcggcc 1260
actgtgggaa aggcctccgt cacctggggt tcttcacgcg gcacagagg actcacaggc 1320
acggggagggt gtagggcgcg ccgaagcgtg ggggtgctgc cctctcgggg agtactgggt 1380
cctgagggag agctgcagtg agaagttgct cttcagcctg gaaaatcaac ctgaattcag 1440
agaagccttc ttagtcctca gagctcccca gtcccccgag aagtttactg ggaaaactgc 1500
caggtgggag aagcagagcc atgggtacgc cggagatggc ggggctctg gagatggcgg 1560
gggctgcgcc ccggcgccgg gcacccctgg gatgtgctga gagtgtgcgc gaccccgag 1620
ccacgtgcca ggccgggctc agaggcgag aagcctgcct ggtgcccaca gccgtctggc 1680
tcagggactc caccctggcc ccgagttgcc gtctgctggg cctttccttc ctggctctgc 1740
accccatgct ggctgcccgg tctggcttcc cttctgtct ctgtcttggg cgaggcagct 1800
gtagcattg cacagaggca aagaccctcc tgcagcctct gcgctgggcc gtagaaacaa 1860
gagcctttgt aataccgaac ctcatccaag gattaggagt ggtggttagg tcaggggccac 1920
ccccagtgt gcaggaacgg cctccacca gctctgttgg tcagagcctg ggtcatgcac 1980
ctggagttgg gagatcaagc tgggtctcag ggcagtgagg tggccatata caccacatcg 2040
catttcgtgg gggaagagggt gacctctttg ttttaaaactt aaggtgtctg cttatccagc 2100
cagaaataaa aatctgccag tgggtgttccc aagggaagac ccccggtggg atggatcggg 2160
acactgccgt gtttcaggtt agccaattat ttttgctttt gcctgttttt ctttaagaat 2220
tgcagccagg ccaggcgtgg tggctcacgc ctgaaatccc agctactcgg gaggtcagg 2280
caggagaatc gtttgaatcc gggaggtgga ggttgtggtg agccaagggt gcgccattgc 2340
actccagcct gggcagtaag agtgaaactc cgtctcaaaa aaaaattcca gccaacagat 2400
gcatatgtaa tgttg 2415
```

&lt;210&gt; 58

&lt;211&gt; 4272

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1907346CB1

&lt;400&gt; 58

```
tcaggctggg tctgtgtcct aaaggggagc tcagtccagc ccagctcccc actgctgcag 60
```

```

tgtgtgtggg cttctccagg aggagggtgg gttctgaaga aaggggtcag aaactcactt 120
cttcctgtag gagtttctcg tccccgtatt cccctggtg cagtggggcg cagaggacca 180
tctttccaca gggtgagatc tgtgcatgtg actgtggcac atgaagggga tgtatagatt 240
ctcagcccta agcagcatat tttggatatt cagtattgac ttctaaagac tcttttatgt 300
tacccaagga agaagtctgg aagaagagga agaggaaaga aaaggagtca gggatggccc 360
ttactcaggt acggttgaca tttagggatg tggccataga attctctcag gaggagtgga 420
aatgcctgga ccctgctcag aggatcttat acagggacgt gatgttggag aactactgga 480
accttgtttc tctgggactg tgtcattttg atatgaatat tatctccatg ttggagggaag 540
ggaaagagcc ctggactgtg aagagctgtg tgaaaatagc aagaaaacca agaacgcggg 600
aatgtgtcaa aggcgtggtc acagatatcc ctccataatg tacaatcaag gatttgctac 660
caaaagagaa gagcagtaca gaagcagtat tccacacagt ggtgttgga agacacgaaa 720
gccttgacat tgaagacttt tcttcaagg aacccagaa aaatgtgcat gattttgagt 780
gtcaatggag agatgacaca ggaaattaca agggagtgtc tatggcccag aaagaaggta 840
aaagagatca acgcgacaga agagacatag aaaacaagct tatgaacaat cagcttggag 900
taagctttca ttctcatctg cctgaactgc agctatttca aggtgagggg aaaatgtatg 960
aatgtaatca agttgagaag tctaccaaca atggttcttc agtgtcacca cttcaacaaa 1020
ttccttctag tgtccaaacc cacaggtcta aaaaatatca tgaacttaac catttttcat 1080
tactcacaca aagacgaaaa gcaaacagtt gtggaaaacc ttataaatgt aatgaatgtg 1140
gcaaggcggt cactcagaat tgaacctta caagtcatag gagaattcat agtggagaga 1200
agccttacia atgcagttag tgcggcaaaa cttttactgt tcttcaaat ctaactattc 1260
atcaggtcat ccatactgga gaaaaacctt acaaagtca tgagtgtggc aaggtcttca 1320
ggcacaattc ataccttgca actcatcggc gaattcatac tggagagaaa cttacaagt 1380
gtaatgagtg tggaaaagcc tttagaggac attcaaacct aactacccat cagttaattc 1440
atactggaga aaaaccgttc aaatgtaatg aatgtggcaa gctcttcact caaaattcac 1500
accttataag tcattggaga attcacactg gagagaaacc ttacaagtgc aatgagtgcg 1560
gcaaagcctt tagtggtcgt tcaagcctag caatccatca gacaatccac actggagaaa 1620
aaccttacia atgtaatgaa tgtggcaaag tcttttaggt caattcatac ctcggaaggc 1680
atcgagagat tcatactggt gagaaacctt acaagtgtaa tgaatgtggc aaagccttca 1740
gtatgcattc aaacctagct acccatcagg tcatccatac tggacaacaaa cctttcaaat 1800
gcaatgaatg cagcaagggt ttcactcaaa attcacaact tgcaaatcat cgaagaatgc 1860
atactggaga gaaaacttac aagtgtaatg agtgtgggaa agccttcagt gttcgttcaa 1920
gtctgactac ccatcaggca atccattctg gagagaaacc ttacaaatgt attgaatgtg 1980
gcaagagcct cactcaaaaa tcacacctta gaagtcatcg ggggaattcat tctggagaga 2040
aaccttacia gtgtaatgaa tgtggtaaag tcttcgctca aacatcacia cttgcaaggc 2100
attggagagt tcatactgga gaaaaacctt acaagtgtaa tgactgtggc agagccttta 2160
gtgatcggtc aagcctaact tttcatcagg caatacatac tggagagaaa cttacaaat 2220
gtcatgaatg cggcaagggt tttaggcaca attcatacct tgcaactcat cggcgaattc 2280
atactggaga gaaaccttac aagtgtaatg agtgtgggaa agcctttagt atgcattcaa 2340
acctaaactac ccataagggtc atccatactg gagagaagcc ttacaaatgt aatcaatgtg 2400
gcaaggtctt cactcagaac tcacaccttg caaatcatca aaggactcac accggagaga 2460
aaccttaccg atgcaatgag tgtgggaaag ccttcagtgt tcttcaagc ctaaccaccc 2520
atcaggcaat ccatactggg aaaaaacctt acaaagttaa tgaatgtggc aaggtcttta 2580
ctcaaaatgc tcacctggca aatcaccgaa gaattcatac tggggagaaa cttacagggt 2640
gtacagagtg tgggaaagcc tttagggtta gatcaagtct aactacccat atggcaatcc 2700
acactggaga aaagcggttac aaatgtaatg agtgtggcaa ggtcttcagg cagagtcca 2760
atcttgcaag tcatcacaga atgcataccg gagagaaacc ttacaaatga gtgtggtgag 2820
gtcattaggt acaattcact cctttcacat cagttaatat cattcttgac agaattcctta 2880
caaatgtagt gacagtggcc aatccctcat gagttgaagc attaatagat atgagaggcc 2940
ataagcaaga gacatcatgt aaacatatgt ggcagagggt ctatccaggc ctgcaggtt 3000
actaggcatc aagatttata tctttgatga aacgaaacaa atgtaatatg catcctgagg 3060
ccattaccca gtgaccgatg gtaagttagg attcctagga ggaataacag tctctggttt 3120
ccctgtttgc ctttgatatt atacactgta gaatactcac aagtccaaat atgctaaaaa 3180
ttatatattt ttaactcaca tacgaaaagg ttgcaggata tttgtaggca gtcagttacc 3240
ttcaccttat gaaatgtttc actgagttat ttaaggtttt ttggaaagcc tactattgcg 3300
tttcaatgtg aactttgaaa tcttattgtg catccttaca caccttccat ggtgctttct 3360
tggaagatc attgggatgg aaggatcatt gattgggtga agatcattga ttaggtgaag 3420

```

```

gattattttct atccaatttg tgaagaagga ggacttttgct tttaaaatta agtatcatct 3480
gaattagcat ttgggagtg cgaaaaacaa tgtaaaacta tgatgtcact caccattctg 3540
ataatgttca ggggtgccttt ctctaccag gagagtactg tggcttagag gaaagaaatg 3600
gtctatcaac tgaacatgaa atggagcagg ccaagacctt aggacattgg gattttttgtg 3660
ggaggagagt aataggtaat tagacactga ttgtgtggta gaaatactgc aggggaaaag 3720
gtcgccctct tatgcatcaa agagcaatac ctgttgttta gcaaagagtg atgaaaaatt 3780
gatcttgttt tgaaattgaa gagagaggcc aggcgcggtg gctcacacct gtaatcccag 3840
cactttggga ggctgaggca ggtggatcac ctgaggtcgg gagttcgaga ccagcctgac 3900
caacatggag aaaccccaat tgtactaaaa atacaaaatt agccgggctt ggtggcagg 3960
gcctgtaatc ccagctactt gggaggctga ggcaggagag tgcttgaacc caggaggcgg 4020
aggttgtggt gagccaagat catgccactg cactccagcc tgggcaacaa gaacaaaact 4080
ccatctaaag aagaaaaaga aaaactgaag agagagtagt tgctcagagt aagagtttaa 4140
tcaacagaag tagagtaaca tattcttttag tagaactcac attatctgct tgaagcacat 4200
tttgctaattg ttttattcag aatgtatcat agatataatg ttttaactaa taaaatagaa 4260
tggtttttct ct 4272

```

&lt;210&gt; 59

&lt;211&gt; 1823

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3041036CB1

&lt;400&gt; 59

```

ggcagagcgg cgagccggtg agtgtggctg cggggctctgc gccaccctc tacctgtgcc 60
gggggcccga acctgcccc ggggtagggc gcggcctcgg ggagacggcg tgcaggccca 120
gcaactgttg tcaccgcagg cctgtcttcc tattcacgtt gggcagcaga gggatggctg 180
ctcagctgct gactgatgag gcaactggaat cagtgcggtt cagggatgtg actgtggact 240
tcaccagga ggagtggcag cagctggagc ctgcccagaa ggacctgtac agggatgtca 300
tgctggagaa ctacaggaac ctggtctcac tggactggga gactagacct gaaatgaaag 360
agttggatcc aaagaatgac atttcggaag acaagctctc cgttggtggg gagggcacgg 420
ggggaccacc gaggaatggt gccaggggtc ctggctcaga aggagtgtgg gaaccaggca 480
gctggccaga gaggcgcgg ggagatgcag gtgcagagtg ggagccattg ggaattcccc 540
aggggaacaa actcttaggg ggctcagtac ccgcagtgtc tgaactgaag gcatttgcca 600
accaaggctg tgtcctggtc ccaccacggc tggacgacct cacagaaaag ggggcctgtc 660
caccgcgaag gcgtggcaag aacttctcca gcaactcaga cctcagtaag cccccatgc 720
cctgcgagga gaagaaaacc tacgactgca gcgagtgtgg caaggccttt agccgaagct 780
cgtccctgat aaagcaccaa aggatccaca cgggagaaaa gccgtttgag tgtgacacct 840
gtgggaagca cttcatcgag cgtcgtccc tcaccatcca ccagcgggtg cacacgggcg 900
agaagcccta tgcctgcggg gactgcggca aggccttcag ccagcgcagt aacctcactg 960
tgcaccagcg cagcacacg ggcgagaagc cgtatgtgtg cgacgtgtgt ggcaaggcct 1020
tccggaagac ttcctctctc acccagcacg agcggatcca caccgggggag aagccctacg 1080
cgtgcgggga ctgcggcaag gccttcagcc agaactgtca cctcatcgtg caccagcgca 1140
cgcacaccgg ggagaagccg tacgtgtgcc ccgagtgcgg gcgagccttc agccagaaca 1200
tgcacctgac cgagcaccag cgcacgcaca ccggggagaa gccgtacgcc tgcaaggagt 1260
gcggaaggc cttcaacaag agtccctcgc tcacctgca ccagaggaac cacaccggcg 1320
agaagcccta cgtgtgcggc gactgcggca aggccttcag ccagagctcc tacctcatcc 1380
agcaccagcg cttccacatc ggctgaagc cgttcagagt cagcagtgct ggcaaggcct 1440
tcagcaagaa cttctcgctc acgcagcacc agcgcattca caccggcgag aagccctacg 1500
agtgctacat ctgcaagaag cacttcacgg ggcgtcgtc cctcatcgtg caccagatcg 1560
tgcacaccgg ggagaagccc tacgtgtgct gcgagtgcgg caaggccttc agccagagcg 1620
cctacctcat cgagcaccag cggatccaca ccggcgagaa gccgtacagg tgcggccagt 1680
gcggaagtc cttcatcaag aactcctccc tcaactgtgca ccagcggatc cacacgggcg 1740
agaagcccta ccgatgcggc gactgcggga agaccttcag ccgcaacacg aacctgacgc 1800

```

gccacctgcg gattcacacc tga

1823

&lt;210&gt; 60

&lt;211&gt; 2253

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3856879CB1

&lt;400&gt; 60

```
gcgggatgcc gttccttcgc gcgtgaggct gcggctctga cgggtgtgtaa cttgtatgtg 60
gaaggaccaa cgtagcattt gcctcagcaa catccttgaa tctgcaccct caccacatc 120
cattcttata agcccatag gctccttcaa tttccgtgat cctcggagtc ccaggagac 180
cagggtgatgg cagcagccag actcctgcc a gtgccggcag gacccaggc caagctgacc 240
ttcgaggatg tggctgtgct cctctcccag gatgaatggg accgcctgtg ccctgctcag 300
aggggcctct acagaaatgt gatgatggaa acctatggga atgtagtctc attgggactt 360
ccaggatcca agcctgacat aatctcccag ctggagcgag gggaagatcc ctgggtcctg 420
gacaggaagg gggctaagaa gagccagggc ctgtggagtg actactcaga caacctcaaa 480
tatgaccaca ctacagcctg tacacaacaa gacagtttat cttgtccatg ggaatgtgaa 540
accaagggag agagtcaaaa tacagacttg agtccgaagc cattaatttc agagcaaaaca 600
gtgattctgg gaaaaacacc cttggggagg attgatcaag aaaataatga aacaaagcaa 660
agcttctgtc tgagtccaaa ctctgttgac caccgtgaag ttcaggctct aagccaaagc 720
atgccactca ctccgcacca ggcagtgcct agtggagaga ggccctacat gtgtgttgag 780
tgtgggaagt gctttggccg gagttccac ctccttcagc atcagcgtat ccacactgga 840
gagaagccct atgtgtgcag tgtatgtggg aaggccttca gccagagctc agtccttagt 900
aaacacagga gaattcacac aggtgagaag ccctatgagt gtaatgagtg tggaaaagcc 960
tttagagtga gctcagatct tgctcagcat cacaagatac atacaggaga gaagcctcac 1020
gaatgtcttg agtgcggaa agccttcact caactctcac atctcattca gcaccagcgg 1080
atccacacgg gagaaaggcc atatgtgtgt ccgttgtgtg ggaaagcctt caaccatagc 1140
actgttctgc ggagccacca gagggtagac actggggaga agcctcacag gtgcaatgag 1200
tgtgggaaaa ccttcagtgt gaagaggaca ctgtctgacg accagaggat ccacaccggg 1260
gagaagccct acacgtgcag cgagtgtggg aaggccttca gcgaccgctc agtcctcatt 1320
cagcaccaca acgtgcacac cggggagaag ccctatgagt gcagtgagtg tgggaagacc 1380
ttcagccacc gctccacact gatgaatcac gagcggatcc acaccgagga aaagccctat 1440
gcatgctacg aatgtgggaa ggccttcgtt cagcactcac acctgatcca gcaccagaga 1500
gtccacactg gggagaagcc ctatgtgtgt ggtgaatgtg ggcacgcctt cagtgcacgc 1560
cggctctctga tccagcatga gagaatccac acagggtgaaa agcccttcca gtgcacagaa 1620
tgtggcaaaag ccttcagcct gaaagcaact ctgattgtgc acctgaggac ccacacgggc 1680
gagaagccat atgagtgcaa tagctgcggg aaggccttca gccagtactc agtgctcatc 1740
cagcaccagc ggatccacac aggcgagaag ccctatgagt gcggggagtg tgggcgtgcc 1800
ttcaaccagc atggccacct aatccagcac cagaaaagtc acagaaaagt gtgacccatg 1860
gctgacacaa gaatccattc tcacagaaac tgcattgtga accacaagca gccttcagcc 1920
caagagaagt ctctgttaac tctataggaa gcttttctt ggcgattcag tgtcacaaaa 1980
taactccaga aagaagcact tagcgtgctg ttctgttggg aaaacttcag agactacctg 2040
ttttattttc ctcaacatct tgaagttagt ttggagagta atcatacaat tgtagagaat 2100
tttggtaaaa aacagccata attctttaac attagtttat ttgaactaag ggaatttaag 2160
gcataagaac cattatccca ataaaatctt acattccaaa taaagttctt tttctaagaa 2220
cattaataaaa aaaaaaaaaa aaaaaaaaaa aaa 2253
```

&lt;210&gt; 61

&lt;211&gt; 2788

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4178665CB1

&lt;400&gt; 61

```
ccgcgccccg ctataatctc tcatttttcta agctcatgtg cctaattttca aattatataa 60
ctttattgaa atattcttta agccactgac ttctagtaag aaccgttttc actttgtttc 120
ctatttttgaa aatgtaaact ttatgttatg ctggctacag gaaaataatt tttgtttgct 180
tttatgtttt ctttcagggt ttttatccag gcataagacc aagaaattat cttcagaaaa 240
ggacattcat gaaatcagtt tatccaaaga gagtataata gaaaaaagta aaactcttcg 300
tctgaaagga tccattttta gaaatgagtg gcagaacaaa agtgagtttg aggggtcaaca 360
gggacttaaa gaaagatcta tcagtcaaaa gaaaatcgct tctaaaaaaa tgtcaactga 420
tagaaaaacgt ccctctttta ctctgaatca gagaattcac aatagtgaga aaagctgtga 480
ctcacacttg gttcaacatg ggaaaataga ttctgatgtg aaacatgatt gtaaagaatg 540
tgggagtact tttaataatg tctatcagct tactctccat cagaaaattc atactgggtga 600
aaaatcctgt aaatgtgaga aatgtgggaa agtttttagt catagctatc aacttactct 660
gcatcagaga ttctatactg gtgagaaacc ctatgaatgt caagaatgtg ggaagacctt 720
tactctttac ccacaactta atcgacatca gaaaattcac actggtaaaa aaccctatat 780
gtgtaagaaa tgtgataagg gtttttttag tagattagaa cttactcaac ataaaaagaat 840
tcatactggg aagaaatctt atgaatgtaa agaattgtga aaagtttttc aacttatttt 900
ctactttaaa gaacatgaga gaattcatac aggtaagaaa ccctatgaat gtaaggagtg 960
tgggaaagct tttagtgtat gcggacaact taccgctcat cagaaaattc atactgggtg 1020
aaaaccctac gaatgtaagg aatgtggaaa gacctttaga cttagttttt accttactga 1080
acacagaaga actcatgcag gtaagaaacc ttatgaatgt aaggagtgtg ggaaatcatt 1140
taatgtgctg ggacagctta atcggcataa aacaatccat actggtataa aaccttttgc 1200
atgtaagggt tgtgagaagg ctttttagtta tagtggtgac ctgagagtac attctagaat 1260
tcatactgga gagaaacat atgaatgtaa ggaatgcggg aaagccttta tgcttcggtc 1320
agtccttact gaacatcag? gacttcatac tgggtgtgaag ccctacgaat gtaagggaatg 1380
tggaagacc tttcgagttc gttctcaaat tagtctacat aagaaaattc atactgatgt 1440
gaagccctac aaatgtgtac gatgtgggaa gacctttaga tttggtttct accttactga 1500
acacagaga attcacactg gtgaaaagcc ctataaatgt aaagaatgtg gaaaggcctt 1560
tattcgtaga gggaatctta aagaacatct gaaaattcat tctggtttaa aaccctatga 1620
ctgtaaagaa tgtgggaagt ccttttagtc gcggtgggcag ttactgaac atcagaaaat 1680
tcatacgggt gtaaaacat acaaatgtaa agaattgtggg aaggccttta gtcgtagtgt 1740
agaccttaga atacatcaaa gaattcatac tgggtgagaaa ccctatgagt gtaaacaatg 1800
tggaagggcc tttagactta attcacacct tactgaacat cagagaattc acactgggtga 1860
gaaaccctat gagtgtgaagg tatgtagaaa ggcctttaga caatattcac atctttatca 1920
acatcagaaa actcataatg taatttaata taagaaaagg tttccatgtc atgctctatt 1980
tatagaatat caaaatattt atggccagaa gttctgtcaa tgtgttgatg tttttttaca 2040
catattaact taataaatgt atgagtcctt aatacctctt agttctcatt aaatttagga 2100
aaattcacac tagaaaataa agctgttaat gtaacagttg tggaaaagtg ttctagcaac 2160
agcatatact tatcatcatt gcctttccac tactctacta tctgtgtgat attagacaaa 2220
atatttgctt cttggtacct cagctgtaaa atgaaacaca cctaaaagtg tggttgtttc 2280
caacatgtat aatacagcaa caactatctg gcccaaactg ctttggatta atattggata 2340
ttactgtttt tattatcatc aacattatta ttagtggatt tcttaatagg aagatgcaat 2400
ggagatgaca aatttggaag aaccactcat cacttacatt tcatgaagta cttctttgat 2460
aaaatctgtt atgggctgaa tgtttgtgtt cccgtaacaa ttctatgtt gaaacacgaa 2520
tcccaagggt atgggtattg aaggtagggc ctttagaggg aaattagggtc atgagggtgg 2580
agccttcacg agtggaatta ctgcctttat aagaagaagc caaagagcca gctagctctt 2640
tcaaccacat gaggttacag caagaagtca gcagtctaca gtgcaaaaaga gggccttcac 2700
cagaacccaa gcattgctggc acctgacct tggactttca gcctccaaa ctgtgagaaa 2760
taaacctcag ttgtttataa aaaaaaaaa 2788
```

&lt;210&gt; 62

&lt;211&gt; 3041

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7493326CB1

&lt;400&gt; 62

gaggaggccg agagatgagg cgggagccca ccaagttctg ggaagttctt tctgacacaa 60  
gtaagataac atctctgccca tgcccacagg tgctgagaac aagagaaatc aagtgaagga 120  
aggaggcgga gtttccaaga cttgggtgtc atcatttctg gggacatcct tgattggaga 180  
ttgaagtttt tgaaccgaaa tttagagctg attcagaaga gacaaatata gaacgtccaa 240  
gttagcagtc atttcccaaa tttaggagac aatgatgcag gccaggaat ccctaacact 300  
ggaggatgtg gctgtggact tcacctggga ggagtggcag ttcttgagcc ctgctcagaa 360  
ggacctgtac cgggacgtga tgttggagaa ctacagcaac ctggtggcag tggggtatca 420  
agccagcaaa ccagatgcac tctccaaatt ggaacgagga gaagaaactt gcacaacaga 480  
agatgaaatc tactctcgaa tctgttctga ttcaggaggt gcatcaggag gtgcatatgc 540  
agaaatcagg aaaattgatg atcctctgca gcatcacttg caaaatcaaa gtattcagaa 600  
gagtgtgaaa cagtgccatg aacagaatat gtttgaaaat attgttaatc agaacaaagg 660  
tcatttctctg ctgaagcaag attgtgatac gtttgactta catgaaaaac ctttaaaatc 720  
aaatttaagt tttgaaaacc agaaaaggag ctctggccta aagaactctg ctgagtttaa 780  
tagagatggg aaatcccttt ttcattgctaa ccataaaciaa ttttatactg aaatgaagtt 840  
tcctgcaatt gcaaaaccta ttaataagtc ccagttcatt aagcaacaga gaactcaaa 900  
catagagaat gcccatgtat gcagtgaatg tgggaaagcc ttcctcaagt tgtctcagtt 960  
tattgatcat cagagagttc acaactggaga aaaacctcat gtatgcagta tgtgtgggaa 1020  
agctttctcc agaaaatcca gactaatgga ccatcagaga actcatacag aactgaaaca 1080  
ttatgaatgc actgaatgtg acaaaacctt cctcaagaaa tcacagctca atatacatca 1140  
gaaaactcat atgggaggga aaccttacac atgtagccaa tgtgggaaag cttcatcaa 1200  
gaagtgtcgg ctcatattatc atcaacgaac tcatacagga gagaaacccc atggtgcag 1260  
tgtatgtggg aaggccttct ctacaaagtt cagtctcact acacatcaga aaactcatac 1320  
aggagaaaaa ccttatatat gtagtgaatg tggaaaaggc ttcattgaga agaggcgtct 1380  
tactgcacat catcgaactc atactgggtga gaaacctttt atatgcaata aatgtgggaa 1440  
aggcttcacc ttgaagaaca gtcttatcac acatcagcaa actcatacag gagagaaatt 1500  
atatacatgt agtgaatgtg gaaaaggctt ttcaatgaag cactgtctca tggtagatca 1560  
acgaactcat actggagaga aaccttataa atgcaatgag tgtggaaagg gcttcgcttt 1620  
gaagagccca ctcatcagac atcagcgaac acatactgga gagaaacctt atgtatgcac 1680  
cgaatgtcga aaaggtttca ccatgaagag tgacctcatt gtacatcagc gaactcatac 1740  
tgcagagaag ccataatatat gcaatgattg tggaaaaggc ttcactgtga agagccgcct 1800  
tattgtgcat cagcgaactc atactggaga aaaaccttat gtatgtggtg agtgtggaaa 1860  
aggctttcca gcaaaagatcc ggctaattggg acatcaacga actcatacag gagagaaacc 1920  
ttatatttgc aatgagtgtg gaaaaggctt cactgagaag agtcatctca atgtacatcg 1980  
gcgactcat acaggagaga aacctatgt atgcagtga tgtggcaaag gcttactggg 2040  
aagagcatgc tcattgcacc atcaggcgaa ctcatactgg ggggagaaac cttatatatg 2100  
caatgaatgt ggaaagggtc tcagcatgaa gagtactctc agtatacatc agcaaaactca 2160  
tactggagag aagccataca aatgcaatga atgtgataaa accttcagga agaagacatg 2220  
cctcatacaa catcagcgat ttcacacagg aaagacttcc tttgcatgta ctgaatgtgg 2280  
aaaatttctt ttgcgcaaaa atgatcttat tacacatcag agaattcaca caggagagaa 2340  
accgtacaaa tgcagtgcact gcgggaaagc cttcactaca aaatcagggc tcaatgttca 2400  
tcaaagaaaa catacaggag agaggcccta tggatgtagt gattgtggga aagcttttgc 2460  
gcacttgtct atccttggtt aacacaagag aattcacagg tagtcatttt gggaaagcct 2520  
cttgccagat gtaggccctt aagatatctg caaagaagag taatttcatg aatgcagact 2580  
acatggttgt ttatttagtg atcagttact tcatgttttg tgtcagagaa aacatgtaca 2640  
aaacatttga gaaaatattt taggacatta tgtctaaaaa ttgtatactg agaaaaatcc 2700  
tatgaatgtg gcagactata aaagcctttg gtgggaagat aaacccctc agaagtgatc 2760  
atagatcatg aataaaactac taattcgtgg aaatgtaata attataggaa cgtcttttgc 2820  
caaaaataaa acctcaacag atttgagaga gtccacactg gagaaaatac ttttctctctg 2880  
gcaggtgaat tgtggcaggg gttttcacgt cacaatgctt ggcctcatcg gtgttgctaa 2940

gaaataattt ccgaaaaaa cttattggac acccttccat gtttcacaga ctttgggaaa 3000  
 caccgagaat ttttgaagga aaaaccttgg aaatgttgtg c 3041

<210> 63  
 <211> 3445  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1553836CB1

<400> 63  
 ctggccggag ccctgggtga aattgttagg cgtggagagg gaggtaggtc ttccagactc 60  
 ggtgctgtac ccgcctagca ggaccagtg ctccataaag gataatagtt tccagtacac 120  
 tatccctcat gatgactcct taagtgggtc atcgtctgca tcttcgtgtg aaccagttag 180  
 tgattttcca gcatctttcc gaaaatctac ctactggatg aagatgagaa gaatcaagcc 240  
 agctgctact tctcatgtcg aagggtcagg tggagtatca gccaaagggg aaaggaaacc 300  
 caggcaggaa gaagatgaag actatcgaga atttcctcag aagaagcata agctttatgg 360  
 gaggaagcaa cggcctaaaa ctcagcccaa tcccaaattc caggcccgtc gtattcggaa 420  
 ggaaccacca gtttatgcag caggcagttt ggaggagcaa tggtagcttag aaatcgttga 480  
 taaaggcagt gtctcctgcc ctacctgcca ggaggtgggg aggaagacca tagagggttt 540  
 aaagaaacac atggaaaact gcaagcagga aatgtttact tgtcatcatt gtgggaaaca 600  
 acttcgttca ctggcaggga tgaagtatca tgtcatggca aatcataata gtttgcccat 660  
 tttgaaagcc ggagatgaaa tagatgagcc aagttagagg gaaaggctcc gaacagttct 720  
 aaagagactg ggaaagctca ggtgcatgct tgagagttgc tccagttagt tcaccagcat 780  
 catgggatat ctctaccatg tcagaaaatg tggcaaaggg gctgcagagc tggaaaagat 840  
 gaccctgaaa tgtcaccact gtgjaaaacc atataggtcg aaggctggac ttgcataatca 900  
 cctgaggtca gagcatgggc ctatatcctt ctttccagag tcaggacagc cagagtgtct 960  
 aaaggagatg aacctagagt caaagagtgg gggccgagtt cagagacgtt ctgccaaagt 1020  
 agctgtatag cacctacagg agctggcctc tgcgaaactg gccaaaggaat ggcccaagag 1080  
 gaagggtgct caggacctgg tacctgatga tcgaaagtta aaatatactc gtccagggtc 1140  
 ccctaccttc agccaggaa tctacataa atggaaagaca gatatacaga aatatcatcg 1200  
 tattcagtggt cctaaccagg gctgtgaggg tgtctacagc agtgtatctg gccttaaagc 1260  
 tcacctgggc tcttgtacat tgggaaactt tgtgggtgga aaatacaaat gtcttctatg 1320  
 tcagaaagaa tttgtgtcag agagtgggtg caagtatcac atcaactccg tccatgctga 1380  
 ggactgggtc gttgttaaacc caacaacaac caaaagcttt gaaaagctga tgaagataaa 1440  
 gcagcggcag caagaagaag aaaagcggag gcagcagcac aggagcagaa ggtctctaag 1500  
 aaggcggcag cagcctggca ttgagcttcc cgagacagag ctgagttcta gagtagggaa 1560  
 ggatcagagg aggaataatg aggaactggg agtgtcagcc tcctgtaagg aaccagagca 1620  
 ggagccagtg ccagcacagt tccagaaagt aaagccccc aagactaatc ataaacgagg 1680  
 aaggaaatag gcagtcagtg taaaagtgtt cctaggaaag cagatgtgat gctgttgtca 1740  
 cggggacctg tgctgggagg actgagtga gattctttca gctgtttgtg taaggctgtg 1800  
 actttctcag ctcttctcct ctgtgtaaat ttcactgtct tcccttctta cattttgatt 1860  
 cccccctccc ctttttagtag gttttcctgc tattcctcgt caagtcctct tgttttttta 1920  
 tcttgcccaa agagctccct ctcaaggcca actataggct cctcttgccc tgtacaaaac 1980  
 taagaaacct ctttggttgt ccttctcttc ctgggggtata gaatgttctt ggaagctcca 2040  
 ttgatttagt agctgtcca tcatctagct tgtgaaacca ttccaaacta atttttttaa 2100  
 accataactg atttgtcatt ttgtatttgt gatataacaa gtctagaagt tagaactgtt 2160  
 gtcattcaca taataagatt actctgtctc cttgggaaaa aaactttatg gaggtgtttt 2220  
 gtcctctcaa tatggtttta agattgaaag taagaaaacg gatttaggat gaaaactcta 2280  
 gaactaccct atgctgttta tactgggaaa tgctttgtac caagtagcag tgactagacc 2340  
 cacagacatg aaaagcaacc ttaggagtaa agtgacccaa acattaaaat gacaggaaag 2400  
 agaaagtaga agcagcaata aatactgcc caactttcct ggagcccag gcctcatcca 2460  
 tagctatgat cacttgccct ctgaagctta tttttgcttc tttggtttta agaattgaga 2520  
 aatatcacat tgcccctgat gttttgacag tctcctagtg ctccgtgtag tgccactaag 2580

```

ggaaaaccaa ggtgcgcatc ccttctccct ggactttacc ttacttggtta gtctacgccc 2640
cactgtttcc acccatcccc ttagccaacc tctgtctttt tgaattttct gagaatattg 2700
tcctatcctc ttttatatat ggagttctct cctctttata tcctgagact ttgacaccag 2760
atgtagatat ttatctggag ctggaaagaa aaattctttt tctgtacctc atgcctatct 2820
ggtaatgttt aatgggttat ttctctttga ggggtggttt ctctggaaca ttggttagag 2880
cagctttgtt gtcgtgtttc tggattctct tccccatttt gcgtaaataat tggctctata 2940
tattcttgcc tattttgtgg catatgccac ataaaaaatg aacctgatat agacaagtac 3000
taccttttca aattctgaaa ggctattacc actttaactc tttgtgctcc tccaaatagc 3060
tttaaaatgt gggcttttgt gaagaccact ttcaaacaag ggagcactga aacctgaatt 3120
ggatactgcc agaataagga gttttgaaac aagttaagga catggtatat gcactctgca 3180
ttttcattgg cagtgtgccc ttaaagccct ttcagtagat gaggggtgtc agggaggaga 3240
aatgaagaag ctatgttaat ttctgggtgag taagacctgg ggaatgtttg gcaatgacaa 3300
aagaaataaa tgactctcag aaagatgttt ccagtgttct ttgacgccga cctgctgcat 3360
gactccttga cactgtatgg gggtaagaag atggctagag atgggggtga gtttgaaata 3420
aattccacat gcagttgtct cagtg 3445

```

&lt;210&gt; 64

&lt;211&gt; 2929

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1908201CB1

&lt;400&gt; 64

```

ccggcctgcc ttcttccctt ccgcccgcgc gccggccgcg cctaagagtt gaccacgccg 60
caatgaagag cttcttacag tgaaagcaaa gtaggtcgag tccactgaaa atgcctttta 120
gggacaaaata ctgtcagact gaccaccatc atcacggatg ctgtgaacca gtttatatcc 180
tggaacctgg agatcctcct ttgttacagc aaccactaca gacatccaaa tctgggtattc 240
aacaaataat tgagtgtctt cgatcaggaa ctaaacaaact taagcatatt ttattaaaag 300
atgtggacac tatttttgaa tgtaagttat gccgcagtct cttcagagga ttaccaaatt 360
taattaccca taaaaaattc tactgcccac caagtctcca gatggatgac aaccttcctg 420
atgtaaatga taaacaaagc caagccataa atgatctcct agaagccata tatccaagtg 480
tggaacaaagc agaataatatt attaagctag aacctataga aactaatcaa aatgcagtat 540
ttcaatatat ttcgaggact gataatccta ttgaagtcac agagtcaagc agtactcctg 600
aacaaccgga agttcagata caggaaacta gcactgaaca gtcaaaaaca gtaccgggta 660
cagatacaga ggtggaaact gtagagcccc ctctgtttga gattgttaca gatgaagttg 720
cacctacatc tgatgaacaa cctcaggagt cgcaggctga cttggaaact tctgacaatt 780
ctgatttttg tcaccagttg atatgttgtc tttgtagaaa agaattcaat tctagacgag 840
gtgttcgccg tcacattcga aaagtacaca agaaaaagat ggaagaacta aaaaagtaca 900
ttgaaacacg aaagaatcca aaccaatcct ctaaaggacg cagtaagaat gttctagttc 960
cattaagtag gagttgtcca gtatgttgta aatcatttgc tacaaaagcg aatgtaagga 1020
ggcattttga tgaagttcat agaggactaa ggagggattc aattactcct gatatagcaa 1080
caaagcctgg gcaacctttg ttctctggatt ctatttctcc taaaaaatct tttaagactc 1140
gaaaacaaaa gtcttcttca aaggctgaat acaatttaac tgcatgcaaa tgcctccttt 1200
gcaagaggaa atatagttca caaataatgc ttaaaagaca tatgcaaatt gtccacaaga 1260
taactctttc tggaacaaac tctaaaagag aaaaaggccc taataatact gccaacagtt 1320
cagaaataaa agttaagtt gaaccagcag attctgtaga atcttcaccc ccttccatta 1380
cccatctctc acagaatgaa ttaaagggaa caaatcattc aatgaaaaa agaacacac 1440
cggcagcaca gaaaaataaa gttaaacaag actctgaaag ccctaaatca actagtccgt 1500
cggctgcagg tggccagcaa aaaaccagaa aaccaaactc ttcagctggc tttgacttta 1560
agcaacttta ctgtaaactt tgtaaacgtc agtttacttc caaacagAAC ttgactaaac 1620
acatcgagtt gcacacagat ggaaataaca ttatgttaa attctacaag tgcctctttt 1680
gcacttatga aactcgtcgg aaacgtgatg tgatacgaca tataactgtg gttcataaaa 1740
agtcactctg ttatcttggg aaaataacag ccagttttaga gatcagagct ataaaaaagc 1800

```

```

ctattgattt tgttctaaat aaagtggcaa aaagaggccc ttcgagggat gaagcaaaac 1860
atagtgatcc aaaacatgat ggcacttcta actctcctag taaaaagtat gaagtagctg 1920
acgtcgggat tgaagtaaaa gtcacaaaaa acttttctct tcacagatgc aataaatgtg 1980
gaaaggcatt tgccaaaaag acttaccttg aacatcataa gaaaactcat aaggc aaatg 2040
cttccaattc acctgaagga aacaaaaacca aaggccgaag tacaagatct aaggctcttg 2100
tctggtgagg aacagttaac agagttttgc ttttttcccc ccatcgaact aaaaaaaaaa 2160
tcatttgacc ataatttata gctggttcca ttttaacacg tttgcttcca tatactctcat 2220
ggcaatggga actgcaagag taatgtgcat attgcattta cctcttcagt gacctttatt 2280
ccagtggcct gggaacaaaa gttaacttca gaacttatct tccacaggac aatgcaatgt 2340
agttgtaggt agatggcaca gggtcagtgc tttcctttta atgttgtaaa atatatacat 2400
ttatattggc ttatgtttac aatagaagtc ttctgtttta ctaacttttg acaggtttta 2460
tttgattcag tgacttagtc tactaattaa tgaattgtag gaaagttaaa tatattagaa 2520
tgaacttgtg aaggcgataa ctataggaaa aaatcttttg aggcaactgta attattgtaa 2580
aaattaatct gtgacggcta aataaaatgc tgcactcaga aaaaaaaatc cccgaattga 2640
attaaaaatg attagcattt atcattttact tttaacatag tgcttccagg caaaggggga 2700
aagttggtgg taaatagaag tttggatttt ttttccttct ccttgggagt gtgtgttagt 2760
tactgttttt tagttttgat ctatggaaac aatgcttatt taggtcccaa atgcttcccc 2820
tctcttttta ataaactttt tgggtgatcat tttaacatag ttaattattat taactaggga 2880
tttggttttt gtggttgtgt gtcgtaagaa aagtaaaatt cataatttg 2929

```

&lt;210&gt; 65

&lt;211&gt; 1923

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2827615CR1

&lt;400&gt; 65

```

ccccggccag gatcgagccc tggcccgggc cctggcccgag ccccggcctc caaggaccgc 60
gccgaaggag ctgcacttcc tcaggatact gtccctctcc cacagaggag ctgaaggagt 120
aggacagaag aactgtcaaa ttctggaatc cttaaagcca tgtccaagga tttggtgaca 180
tttgggggat tggctgtaaa tttctctcaa gaggaatggg aatggctgaa ccctgctcag 240
aggaatttgt acaggaaagt gatgttggag aactacagga gcttggatc attggcagga 300
gtttctgttt ctaagccaga tgtgatctca ttattggagc aaggaaaaga gccctggatg 360
gtgaagaagg agggaaacaag aggcccatgc cctgattggg agtatgtatt taaaaacagt 420
gaattttcat caaagcaaga gacatatgaa gaatcatcca aagttgtgac agtgggagca 480
agacatctta gttatagcct tgactatccc agtttgagag aagactgtca aagtgaggac 540
tggtataaga accagctggg aagtcaagag gtacatctta gtcaattaat catcactcat 600
aaagaaatcc ttccagaagt tcaaaaataa gaataaaca aatcttggca aacattccac 660
caggatacaa tctttgatat acaacagagt tttcccacca aagaaaaagc acataagcat 720
gaaccacaaa agaaaagtta ccgaaaaaaa tctgttgaaa tgaaacatag gaaagtctat 780
gtagaaaaga aacttttgaa atgtaatgat tgtgagaaag tcttcaacca gagctcatcc 840
cttactcttc atcagagaat tcatactgga gagaaaccct atgcatgtgt tgaatgtggg 900
aaaacgttca gccagagtgc aaacttggcg caacataaga gaatacatac tggagagaaa 960
ccctatgaat gtaaagaatg taggaaagcc ttcagccaga atgcacacct ggcccaacat 1020
cagagagttc atactggaga gaaaccttat cagtgtaaag aatgtaaaaa agccttcagc 1080
cagattgcac acctgactca gcatcagaga gttcatactg gagagagacc tttcgaatgt 1140
attgaatgtg gaaaggcctt tagtaatggt tcatttcttg ctacgcatca gagaattcat 1200
acaggagaga aaccttatgt gtgtaatgtg tgtgggaaag cctttagcca tcgtggatac 1260
ctaattgtac atcagagaat tcatactgga gagagaccct acgaatgtaa ggaatgtagg 1320
aaagccttca gccagtatgc acaccttgct caacatcaga gagttcatac tggagaaaaa 1380
ccttatgaat gtaaagtatg taggaaagcc ttcagccaaa ctgcatacct tgatcaacat 1440
cagagggttc atactggaga gaaaccctat gaatgtattg aatgtgggaa ggcctttagc 1500
aatagttcat cacttgcaca acatcagaga agtcatactg gagaaaaacc ctatatgtgt 1560

```

```

aaggaatgta ggaaaacatt. tagccagaat gcaggccttg ctcaacatca gagaattcat 1620
actggagaga aaccttatga atgtaatggt tgtgggaaag cgtttagcta tagtggatct 1680
cttactctac atcagagaat tcatactgga gaaagaccct atgaatgtaa agattgcagg 1740
aaatctttca ggcagcgtgc acatcttgct catcatgaga gaattcatac tatggagtca 1800
ttcttgactc tttcctctcc ctcaccctcc acatcaaatac agttgccaaag acctgtagg 1860
ttcatctcct gaataatttct ggaatccacc tcttgaatcc atttccatcc catcatcctt 1920
gtc 1923

```

&lt;210&gt; 66

&lt;211&gt; 5601

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4304550CB1

&lt;400&gt; 66

```

actccctccc cgcgggggcgc gcagctcgcg ggtcttttga caccaccggt cctgagtcgg 60
cggactgcca ttttcattaa gaactgccac ttagaggtag caaaataaag ggtattttgct 120
acctttaata cttgccagtt cagggttgag gcacaggcag cagcaagaat ggaaagaaat 180
gttcttataa cattttcaca ggaaatgtcc cagttaattt tgaatgaaat gccaaaagct 240
gaatattcca gtttattcaa tgattttgtt gaatctgaat tttttttgat tgatggggat 300
tcattactta tcacatgtat ctgtgagata tcatttaagc ctgggcagaa cctccatttc 360
ttctatctgg ttgaacgcta tcttgtggat cttatttagc aaggaggaca attcaccata 420
gttttcttca aggatgccga gtatgcgtat ttcaacttcc ctgaacttct ttctttgaga 480
actgctttaa tccttcatct tcagaagaat accaccattg atgttcgaac aacatttttcg 540
agatgcttat caaaagagtg gggaaagtct ttggaagaga gttaccata tttcctgata 600
gttgacagc aaggcctgaa cgatctacaa acacagcttt tcaacttttt aatcattcat 660
tcttgggcaa ggaaggtcaa cgttgtactt tcctcagggc aagaatctga tgttctttgc 720
ctttatgcat accttcttcc aagcatgtac agacaccaga ttttttcttg gaagaataag 780
cagaacatta aagatgctta tacaacctg ctttaaccagt tggaaagatt taagctttca 840
gcatttagcac ctcttttttg aagtttaaaa tggaataata ttacggaaga ggcacacaag 900
actgtatctc tgcttacaca agtctggcca gaaggatctg acattcggcg tgtcttttgc 960
gttacttcat gctcattatc tttgagaatg taccatcgct ttttaggaaa cagagagccc 1020
tcctctggtc aggaaactga gatccaacag gtgaacagta attgcttaac cctgcaggag 1080
atggaagatt tgtgtaaact gcattgtctc actgtgggtt ttctactcca tctgcctctt 1140
tctcaaagag cttgtgctag agtcatcact tccattggg ctgaggacat gaagccttta 1200
ttacaaatga aaaagtgggtg tgaatatttc atcttaagaa atatacatac ttttgaattt 1260
tggaatctga atttaattca cctttctgac ttaaatgatg agcttttgtt gaagaatatt 1320
gctttttact atgaaaatga aaatgtaaaa ggcctacatt tgaatttggg agataccatt 1380
atgaaagatt atgaatatct ctggaatacc atatcaaagt tggtcagaga ctttgagggt 1440
ggacagccat ttctcttgag aacaacaaaa gtttgttttc ttgaaaagaa accatcacca 1500
atcaaagaca gtcctaatga aatgggtgcc aatttgggtt ttattccaac gtcacttttt 1560
gtggttgata aatttgctgg agatattttg aaagatttgc cttttctaaa gagtgatgat 1620
cctattgtta cttcactggt taaacaaaag gaatttgatg aacttgtgca ctggcattct 1680
cataaacccc tgagtgatga ttatgacagg tccagggtgc agtttgatga aaaatctaga 1740
gaccctcgtg ttcttagatc tgtgcaaaaag tatcatgttt tccaacgggt ttatgggaat 1800
tcattagaaa cagtctcttc gaaaatcacc gtgactcaaa ctattaagtc aaagaaggat 1860
tttagtgggc ccaagagcaa aaaggcacac gagaccaagg ctgaaataat tgctagagag 1920
aataagaaaa ggttattttgc cagggaagaa caaaaggaag agcaaaagtg gaatgctttg 1980
tcattttcta ttgaagagca attgaaagaa aattttacact ctggaataaa gagcctggaa 2040
gattttttga aatcctgtaa aagtagctgt gtgaaacttc aggttgaaat ggtgggggta 2100
actgcttgct tgaaagcctg gaaagaacat tgccgaagtg aagaaggtaa aaccacgaaa 2160
gatttaagta tagctgttca ggtgatgaaa aggatccact ccttgatgga aaaatactca 2220
gaactttttac aagaagatga tcggcaactc atagccagat gccttaagta tttaggattt 2280

```

```

gatgagttgg caagttcttt acatccagcc caggatgcag aaaatgatgt aaaagtgaag 2340
aaaaggaata aatattcagt tggcattggg ccagctcggg tccaactgca atacatgggc 2400
cattatttga tacgagatga gagaaaagac ccagatccca ggggtccagga ttttattccc 2460
gacacatggc agcgagagct ccttgatggt gtggataaga atgagtcagc agtgattggt 2520
gccccaacgt cctcaggcaa aacctatgcc tcctactact gtatggagaa agtgctgaag 2580
gagagcgacg acggggtggg cgtgtacggt gcacccacaa aggcccttgt taatcaagtg 2640
gcagcaactg ttcagaatcg ttttacgaaa aatctgccaa gtgggtgaagt tctctgtggg 2700
gttttcacca gggagtatcg tcatgatgcc ttaaaactgtc aggtacttat tacagtgcct 2760
gcctgctttg aaattctgct gcttgctcct catcgccaaa actgggtgaa aaagatcaga 2820
tatgttatat ttgatgaggt tcattgtcct ggtggagaaa ttggagcaga aatctgggaa 2880
catctccttg tcatgatccg atgtcccttt ttggctcttt cagctaccat aagtaatcct 2940
gaacatctca ccgagtggct acaatcggta aaatgggtact ggaaacaaga agacaaaata 3000
attgaaaata ataccgcttc taaaagacat gtgggtcgtc aggccggctt tcccaaagac 3060
tacttgcaag taaaacaatc gtataaagtt agacttgtgc tctatggaga gaggtataat 3120
gatctagaga agcatgtatg ttcaataaaa catgggtgaca ttcatttttga tcattttcac 3180
ccatgtgctg cactaacaac agatcatatt gaaaggtatg gattccctcc tgatcttacc 3240
ctttcacctc gagaaagcat ccagctgtat gatgccatgt ttcaaatttg gaaaagttgg 3300
cctcggggccc aggaactgtg ccagaaaaac ttcattcatt ttaacaataa attagtcatt 3360
aaaaagatgg atgctaggaa atatgaagag agtctaaagg cagaattaac aagttggatt 3420
aaaaatggca acgtagagca ggccagaatg gtacttcaga atcttagtcc tgaagcagat 3480
ttgagtcagc aaaacatgat caccatgttt ccacttctag ttgaaaaact aaggaaaaatg 3540
gagaagttac ctgcactatt ttttttattc aagttaggag ctgtagaaaa cgcagctgaa 3600
agtgtgagca ctttcctaaa gaaaaagcag gagacaaaaa ggccctccca agctgataaa 3660
gaagcccatg tcatggctaa caaacttcga aaagttaaaa aatccataga gaaacaaaag 3720
atcatagatg aaaagagcca gaaaaaaacc agaaatgtgg atcaaagcct aatacatgaa 3780
gctgaacatg ataacttagt gaagtgtcta gagaagaacc tggaaatccc acaggactgc 3840
acatatgctg atcaaaaagc agtggacact gagactttgc agagggtatt tggctgagta 3900
aaatttgaag gaaaaggtga agaattgaaa gccttggcag aaaggggtat tggatatcat 3960
cacagtgcta tgagtttcaa agaaaaacaa ttagttgaaa tcctctttag aaaaggatat 4020
cttaggggtg tgacagctac tggaacactt gctttagggt tcaacatgcc ttgtaaatct 4080
gtgggttttg ctcaaaactc agtctatctg gatgcgttga attatagaca gatgtctggc 4140
cgtgctggaa gaagagggtca agacctgatg ggagatgtat atttctttga tattccattc 4200
cccaaaatag gaaaactcat aaaatccaat gttcctgagc tgagaggaca ctccctctc 4260
agcataaccc tggctcctgcg actcatgctg ctggcttcca agggagatga ccagaggat 4320
gccaaggcaa aggtgctatc agtgcataag cattcattgc tgtccttcaa gcaaccaga 4380
gtcatggaca tgttaaaact ttacttcctg ttttctttgc agttcctggg gaaagagggc 4440
tatttagatc aagaaggtaa tcctatgggg tttgctggac ttgtatcaca tttgcattat 4500
catgaacctt ctaatcttgt ttttgtcagt tttcttgtaa atggactctt ccatgatctc 4560
tgtcagccaa ccaggaaagg ctcaaaacat ttttctcaag acgttatgga aaagctagta 4620
ttagtatggg cacatctctt tggagaaga tattttccac caaagttcca agatgcacac 4680
ttcaggtttt atcaatcaaa ggtgttcctt gatgatctcc ctgaggattt tagtgatgct 4740
ttagatgaat ataacatgaa aattatggag gactttacca ctttcctacg aattgtttcc 4800
aaactggctg atatgaatca ggaatatcaa ctccattgt caaaaatcaa attcacaggt 4860
aaagaatgtg aagactctca actcgtatct catttgatga gctgcaagga aggaagagta 4920
gcaatttcaa catttgtttg tctgtctggg aactttgatg atgatttgct tcgactagaa 4980
actccaaacc atgttactct aggcacaatc ggtgtcaatc gctctcaggc tccagtgcgt 5040
ttgtcacaga aatttgataa ccgaggaagg aaaatgtcgc ttaatgccta tgcactggat 5100
ttctacaaac atgggttcct gataggatta gtccaggata acaggatgaa tgaaggagat 5160
gcttattatt tgttgaagga ttttgactc accattaaat ctatcagtgt ttccttgctg 5220
gagctatgtg aaaatgaaga cgacaacggt gtcttagcct ttgaacaact gagtacaact 5280
ttttgggaaa agttaaacaa agtctaaaaa caaagtctat gcaaaccact taaaaataat 5340
tccatagtag tttttcaggt cacgtttttg attcttatgc ttcttgccag aaatacatta 5400
tgataaagtg gaaatacatt acgatgaagt ggaaagagca aacactttgg aatcaaacag 5460
agttgcaatc aaacctgcca tgttctgtca tgaatactca caaattatgt agtataacct 5520
aatcttgggt tctttttata actgagtaat aatgggtaca tgtcagggga tccactagtt 5580
taagcgccgc ccccggtgca g

```

<210> 67  
<211> 2802  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7473738CB1

<400> 67  
gacccccctac gtcacacagg cgcgctcctcg cctttccccc cccatctcct cggctccgtg 60  
gacgtgttcg cttcttctcc ggaccgggtc tatgtcccg tttcccgccg tcgcgggcag 120  
ggcgccaagg cggcaggagg aaggcgagcg gccaatagag ctccaggag agcgggccgtc 180  
agctgttcgc atcgctgaca gagaagagaa aggatgcaca tcacaggagg gaggaacaac 240  
tccaacgttt cctattcaga aacaaagaaa aaagctcatt caagctgtga gggacaactc 300  
gttcctcatc gttactggaa atacaggaa tggtaaaaa actcaactcc ctaaatacct 360  
ttatgaagca ggggttttcac aacatgggtat gattggtgtg actcaaccac gaaaagtagc 420  
tgctatatca gttgctcaga gagtagctga agaaatgaaa tgcactttgg gatctaaagt 480  
aggataccaa gttcgttttg atgattgcag ttctaaggag acagcaatca aatatatgac 540  
tgatggatgt ttactgaaac atattctggg agacccaaat cttaccaaat tcagtgtcat 600  
tattttggat gaagcccatg aaagaactct aactacagat atcttatttg gtttactgaa 660  
gaagctatth caggagaagt ctccaatag gaaggagcat ttaacaagtg gtggtacatg 720  
tcacgcaact atggaattag ccaagctctc tgcattcttt ggaaattgcc ccatatttga 780  
tatacctgga agactctatc cagtcagaga gaaattctgc aatttgattg gtccacgaga 840  
cagagaaaat actgctgata ttcaagcagat tgtgaaagtc accatggata tccatttgaa 900  
tgaaatggct ggagacatct tgggttttct gactggccag tttgaaatag aaaaaagttg 960  
tgagttactt tttcagatgg cagagtctgt tgattatgat tatgatgttc aagataccac 1020  
cctcgatggc ttgttaatat tgccgtgtta tggatcaatg acaacagatc aacagaggag 1080  
gatatttttg ccaccaccac ctggaattag aaaatgtgtc atatccacca atatttctgc 1140  
aacgtctttg acaatagatg gaatcagata tgtggtagat ggtggctttg tgaagcagtt 1200  
aaatcacaa cccagattag ggttggacat cctggagggtg gttccaattt caaagagcga 1260  
ggcattacag cgaagtggcc gagctggcag gacttcttca ggaaaatgct ttcggatcta 1320  
tagtaaagat ttttggaacc agtgtatgcc tgaccatgtg atccctgaaa ttaagagaa 1380  
tagtttgaca tctgtagttc tgaccttaaa gtgccttgcc atacacgatg tcataagggtt 1440  
tccctatttg gatccaccta atgagagact tattttagaa gctcttaaac aactttacca 1500  
gtgtgatgct attgacagga gtggccatgt caccagattg ggtttgtcta tgggtggagt 1560  
tcctttgcct ccacatctga catgtgcagt aataaaagct gcttccctgg attgtgaaga 1620  
tctactactt ccaatagcag caatgtttgtc tgtggaaaac gtcttcatta gacctgttga 1680  
tccagagtac cagaaggaag cagaacagag acatcgagaa ttggcagcta aagctggagg 1740  
atttaattgac tttgcaactt tagctgtcat ctttgaacaa tgcaaatcaa gtggagctcc 1800  
agcttcatgg tgccaaaaac actggattca ttggagggtgc ttattttctg catttctgtg 1860  
ggaagctcaa cttcgagaac taatcaggaa gcttaaacag caaagtgatt tcccaaaaaga 1920  
gacttttgaa ggccctaaac atgaagtact acgaagatgt ctttgtgcgg gctacttcaa 1980  
aatgttgtct cgaagatctg ttggcagaac attttgcaca atggatgggc gtggaagtcc 2040  
agttcatatc catccttctc cagcacttca tgaacaggaa accaaacttg aatggataat 2100  
ttttcatgaa gtattagtta ccaccaagt ctatgcaaga attgtgtgcc caatccgtta 2160  
tgaatgggtg agagacttac taccgaagt gcatgaactt aatgcgcatg atttgagcag 2220  
tgtggctaga cgtgaaatga gagaagatgc aagaagaaaa tggacaaata aggaaaatgt 2280  
gaaacagcta aaggatggaa tatcaaaaga agttctaaag aaaatgcaaa gaagaaatga 2340  
tgacaaatcc atatcagatg cacgggctcg cttccttgag agaaagcaac agagaatcca 2400  
ggaccacagt gacacattaa aggaaactgg ctaagcagtg gactatcaa ttcagggaagc 2460  
aggaaagcag ccaggaaatg tgcttctgtt ttgccagtta tttccgacag tactatcaga 2520  
aggaaagtgg cagcagctgt tcctagcctg tgagccgaaa gcaaatcaga atttataaat 2580  
cacatctcat caatactaac aaatgacatt ttgaaaggaa agattgggtt gcatagtcac 2640  
agtacatgaa tccaatcaaa gatgatacaa tatttccctc acttctgttt tgctggcctt 2700  
aactgggtatc aaacagtggtc actgagatat tttcaaagaa cactgaattg tatttaatac 2760

gcgtgtattc catttgcat gaagcattaa agattatttt cc

2802

&lt;210&gt; 68

&lt;211&gt; 2157

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4447743CB1

&lt;400&gt; 68

gggcttcagc tctctgcgtt ctcggtccg ggaggcctcg gtgattcagc cacagcctct 60  
gcctcccgtt gctctgtgac ctgaggggat tggacaattt gtagctaaga ctcccggata 120  
ccctgaagtc gggaaatgga actcgtaaca ttcaggggatg tggccataga attctcccct 180  
gaagagtgga aatgtctgga ccctgcccag cagaatttgt atagagatgt gatgttgag 240  
aactacagga acctggtctc cctgggtttt gtgatctcta acccagacct ggtcacctgt 300  
ctggagcaaa taaaagagcc ctgcaatttg aagatacatg agacagcagc caaaccccca 360  
gctatatgtt ctcttttcag ccaagacctt tcaccagtgc aggggataga agattcattc 420  
cacaaactta tactgaaaag atacgagaaa tgtggacatg agaatttaca attaagaaaa 480  
ggctgtaaac gtgtgaatga gtgtaagggtg cagaaaggag ttaataatgg agtttaccag 540  
tgcttgtcaa ctaccagag caaaatattt caatgtaata catgtgttaa agtttttagt 600  
aaattttcaa attcaaaca acataagata agacatactg gagagaaacc ctttaaattgt 660  
acagaatgtg gcagatcggt ttacatgtca cacctaactc aacatacagg aattcatgct 720  
ggagagaaac cctacaaatg tgaaaaatgt ggcaaagcct ttaatagggtc cacatcactt 780  
agtaaacata agagaattca tactggagag aaaccctaca catgtgaaga atgtggcaaa 840  
gccttttagac ggtccacagt tctgaacgaa cataagaaaa ttcatactgg agagaaacc 900  
tacaaatgtg aagaatgtgg caaagccttt acaagggtcca caacactgaa tgaacacaag 960  
aaaattcata ctggagagaa accctacaaa tgtaaagaat gtggcaaagc ctttagatgg 1020  
tccacaagcc tgaatgaaca taagaatatt catactggag agaaacccta caaatgtaaa 1080  
gaatgtggca aagccttttag acagtccagg agcctgaatg aacataaaaa tattcact 1140  
ggcgaaaaac cctacacatg tgaaaaatgt ggcaaagcct ttaaccaatc ctcaagtctt 1200  
attatacaca ggagcattca ttctgaacaa aaactttaca aatgtgaaga atgtggcaaa 1260  
gcctttactt ggtcctcatc ccttaataaa cataagagaa ttcatactgg agagaaacc 1320  
tacacatgtg aagaatgtgg caaagccttt tataggtcct cacaccttgc taaacataag 1380  
agaattcata ctggagagaa accctacacg tgcaagaat gtggcaaagc ttttaaccaa 1440  
tcctcaactc ttatattaca caagagaatc cattctgggc aaaaaccta caaatgtgaa 1500  
gaatgtggca aagcctttac acggtccaca aactgaacg aacataagaa aattcact 1560  
ggcgagaaac cctacaaatg tgaagaatgt ggcaaagcct tcatatgggtc cgcaagcctg 1620  
aatgaacata agaattattca tactggagag aaaccctaca aatgtaaga atgtggcaaa 1680  
gcttttaacc aatcctcagg ccttattata cacaggagca ttcattctga aaaaaactt 1740  
tacaaatgtg aagaatgtgg caaagccttt actcggtcca cagccctgaa tgaacataag 1800  
aaaattcatt ctggagagaa accctacaaa tgcaagaat gtggcaaagc ctataactta 1860  
tcctcaacc ttactaaaca taagagaatt catactggag agaaaccctt cacatgtgaa 1920  
gaatgtggca aagccttcaa ttggtcctca tcccttacta aacataagat aattcact 1980  
ggagagaaat cctacaaatg tgaagaatgt ggcaaagcct ttaatcggtc ctcaaccctt 2040  
actgtacaca agcgaattca tactggcaag gaacatagtt gaatgacatt tctagtaatc 2100  
tctaattcca gtgtctttac acagcaaata aattggagaa tattgctccc atataaa 2157

&lt;210&gt; 69

&lt;211&gt; 2104

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7497554CB1

&lt;400&gt; 69

```
gcgcctgccc cggtgggtgg gaagcaggac tcgggcgctc atgcagcgag cgggcggcgc 60
tcggggcgct agttcctagc agctgggcca gcgtaggggc gcaggtcgga tccggcagag 120
gagggcgag gagacgcag ggggaggggtg gagagacccg cgagccgcag tctcagcctc 180
gtccgacgcg cctccgcctc tcccgggccc ggcccgggtg gcgctcagag cttgagggcg 240
ccggctgctc cctcggtagc gggggcaagc ggaggcaggg gtgtgggccc ctaaaatgag 300
tgaaaggaga agatctgcag tcgccctgag ctgcgcagca catgccttct ccgttgaagc 360
cttgatcggc tcaataaaaa aacggaaact gcgagactgg gaggagaagg ggctggacct 420
gtctatggag gcgctgagcc ccgcgggccc actcggagac acggaggacg cggcggcaca 480
cggcctggag cctcaccggg attctgagca gagcactggg tcagattctg aggtccctcac 540
tgagcggact tctgtctcct tcagtactca cactgacctg gcctctgggtg ctgcaggccc 600
tgtgcctgct gccatgtctt ccatggagga gattcagggtg gagctgcaat gtgctgacct 660
ctggaagcgg ttccatgata ttggaactga aatgatcatc accaaagcag gcaggaggat 720
gtttcctgcc atgagagtga aaatcactgg cctagatcca aatcagcagt actacatagc 780
aatggacatt gtgcctgtgg acaataaaa atacagatat gtgtatcata gctccaagtg 840
gatgggtggc ggcaatgctg attcccctgt gcccacaaga gtttatatac accctgattc 900
tctagcttct ggagacacct ggatgagaca ggtggtcagt ttgacaaac tcaagcttac 960
caacaatgag ttggatgac aaggacatat cattctgcac tctatgcaca aataccagcc 1020
tcgagttcat gtgattcgca aagacttcag cagtgcacct tcacccacta agcctgttcc 1080
tgttggggat ggggtgaaaa cgttcaactt tcctgagact gtgttcacca cagttacggc 1140
ctatcagaat cagcagatta ccagattaaa aattgaccga aacccttttg ctaaaggatt 1200
cagagattct gggagaaaca gaactggact tgaagccatc atggagacat atgcattctg 1260
gagacctcct gtgcgcacac tcaccttcga agacttcacc accatgcaga agcagcaagg 1320
aggcagcaca ggcacttccc caaccacctc cagcactggg acaccatccc cttcggcttc 1380
ttctcatctt ttatctccat cctgttctcc tccaactttt catctggccc ccaacacttt 1440
caatgtgggc tgccgagaaa gccagctgtg taatctaaac ctctctgatt atccaccatg 1500
tgcccgaagc aacatggctg ccttgagag ctaccagggg ctgagtgaac gtggctacaa 1560
caggcttcag agtggcacca cttcagccac tcagccctct gaaaccttca tgcctcagag 1620
gactccatcc ctgatctcag gaataccaac tcctccctcg ttgcctggca acagcaagat 1680
ggaagcctac ggtggccagc tggggctcct tcccacttcc cagtttcagt atgtcatgca 1740
ggcaggcaat gctgcctcca gctcctcatc accacacatg ttcgggggca gccacatgca 1800
gcagagctcc tacaatgcct tctcccttca caaccttac aacctgtatg gatacaattt 1860
ccccacttcc cctaggctag ctgcaagccc ggaaaaactg agcgcctctc aaagcacttt 1920
actctgttct tctccttcca acggggcctt tggagagagg cagtacctgc cgtcagggat 1980
ggagcacagc atgcacatga ttagcccttc acccaataac caacaggcaa ccaacaattg 2040
tgatggccgg cagtatgggg cagttccagg ctccctctcc cagatgtccg tgcacatggg 2100
ttaa 2104
```

&lt;210&gt; 70

&lt;211&gt; 1451

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7475843CB1

&lt;400&gt; 70

```
tttcagtctc ttctataagt aagaaccagt tctctctgtc tctcttgtgc ctggaatcag 60
attaagaact ctgccatga ccacttgata cgtatgtgtt ttttctttgc cagggactgt 120
tgatattcag ggatgtggcc atagaattct ctccggagga gtggagctat ctggacctg 180
ctcagcagaa tctgtatagg gacgtgatgt tagaaaacta cagaaacctg gtctccctgg 240
gtattgtctg ctctaagcca gacctgatca cctgtctaga gcaaaggaaat gagccctgga 300
atgtgaagaa acatgagaca gtagccagac acccagctgt ttctttctcat ttcacccaag 360
```

```

acctcttgcc agagcatggt ataaaagatt catttcaaaa agtgatactg agaagatatg 420
gaagctatgg cattgagaat ttacaattaa agaaagattg ggaaagtgtg ggtgaatcta 480
agggtgcagaa agaattgtgt aatggactta accaatcttt atcaactaca cataccaaaa 540
tctttcaatt taataaatgt gtgaaagtct ttagtaaata gtcaaatacta aatagacata 600
agataagaca tactggagag atatcttcca actgtaaaga atgtgacaat tcctttttaca 660
tatectcagt tctaactcca cttcagagaa tccacactgc agagaaatcc tacaagtgtg 720
aacaatgtgg gaaagccttt aggcactgct catgctttct tgaacatgag acaattcata 780
atgaagagaa acattacaaa tgtaaagaat gtggaaaagt ctttaaatcc ttcacaagcc 840
tttctaatac cattataatt catactggaa agaaactcta taaatgtgaa gaatgtggca 900
aagcttttaa ccacagttca aacctggcca aacataagaa aattcacact ggacagaaac 960
cccataaatg tgaagaatgt ggcaaagcct ttaactgggt ctcataccta actctacata 1020
aaagaattca tactggagag aaaccttaca aatgtgatga atgtggcaaa gcttttaacc 1080
agtgttcaaa cctcactaaa cataagagaa ttcatactgg agagaaaccc taaaaatgtg 1140
aagaatgtgg caaagctttt aaccgggtgct cacaccttac tgaacataaa agaattcata 1200
ctggagagaa gccctataaa tgtgaagaat gtggtaaagt ctttatatct tgttcaagcc 1260
tttcaaacca taagagaatt catacaagag agaaatgcta caaatctgaa gaatgtggca 1320
aaacctttta ccaactgctca gacctcaatg tacctgagaa aattcatacc tgagaaaaat 1380
cctacaaatg taaaaaatgt ggcaaagcct ttaatacctg ctcagtgtct actcaggacc 1440
agagttcata t 1451

```

&lt;210&gt; 71

&lt;211&gt; 1609

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6319550CB1

&lt;400&gt; 71

```

agggaggctg cgtgtgccgg ggctaggggc tggaaagtcct ggctctagtt gcacctcgga 60
aggaaaaggc aaacagagga gggaaggcgt cttaggactg cctggatcca gagcactttc 120
cacggcctct acaggcctgt gtcgctatgg gttccccgc cgccccggag ggagcgctgg 180
gctacgtccg cgagttcact cgccactcct ccgacgtgct gggcaacctc aacgagctgc 240
gcctgcgcgg gatcctcact gacgtcacgc tgctgggttg cgggcaaccc ctcagagcac 300
acaaggcagt tctcatcgcc tgcagtggct tcttctattc aattttccgg ggccgtgcgg 360
gagtcggggg ggacgtgctc tctctgcccg ggggtcccga agcgagaggc ttcgccccctc 420
tattggactt catgtacact tcgcgcctgc gcctctctcc agccactgca ccagcagtcc 480
tagcggccgc cacctatttg cagatggagc acgtgggtcca ggcatgccac cgcttcatcc 540
aggccagcta tgaacctctg ggcatctccc tgcgccccct ggaagcagaa ccccaaacac 600
ccccaacggc ccctccacca gtagtccca ggcgctccga aggacacca gaccaccta 660
ctgaatctcg aagctgcagt caaggcccc ccagtccagc cagccctgac cccaaggcct 720
gcaactggaa aaagtacaag tacatcgtgc taaactctca ggctcccaa gcaggagacc 780
tggtcgggga gagaagttct ggtcaacctt gcccacaagc caggctcccc agtggagacc 840
aggcctccag cagcagcagc agcagcagca gcagcagcag tgaagaagga cccattcctg 900
gtccccagag caggctctct ccaactgctg ccactgtgca gttcaaatgt ggggctccag 960
ccagtacccc ctacctctc acatcccagg ctcaagacac ctctggatca ccctctgaac 1020
gggctcgtcc actaccggga agtgaatttt tcagctgcca gaactgtgag gctgtggcag 1080
ggtgctcatc ggggctggac tccttggttc ctggggacga agacaaaccc tataagtgtc 1140
agctgtgccg gtcttcgttc cgctacaagg gcaaccttgc cagtcaccgt acagtgcaca 1200
caggggaaaa gccttaccac tgetcaatct gcggagcccg ttttaaccgg ccagcaaacc 1260
tgaaaacgca cagccgcac cttcggggag agaagccgta taagtgtgag acgtgctggc 1320
cgcgctttgt acaggtggca catctgcggg cgacagtgct gatccacacc ggggagagac 1380
cctacccttg ccctacctgc ggaaccgct tccgccacct gcagaccctc aagagccacg 1440
ttcgcatcca caccggagag aagccttacc actgcgaccc ctgtggcctg catttcggc 1500
acaagagtca actgcggctg catctgcgcc agaaacacgg agctgctacc aacaccaaag 1560

```

tgcactacca cattctcggg'gggcctagc tgagcgcagg cccaggccc

1609

&lt;210&gt; 72

&lt;211&gt; 2840

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7510064CB1

&lt;400&gt; 72

ggagcggcga ctggcgagcc atggcgctgg ggctgcagcg cgcaaggccg gccctttcct 60  
gtggagtcac ctccaccgccc tgccgaccca ctcgtaactc gcaccggggt cctggctgca 120  
ccgcatcccc tcctgcaccc cctggatggc ccttcagcca acgggggcct gggcgatggg 180  
cgaccacgga gctgcgcaag gaaaagtccc gggatgcggc ccgcagccgg cgcagccagg 240  
agaccgaggt gctgtaccag ctggctcaca cgctgccctt cggccgcggc gtcagcgcgc 300  
acctggacaa ggctctctatc atgcgcctca ccatcagcta cctgcgcatg caccgcctct 360  
gcgccgcagg ggagtggaa caggtgggag cagggggaga accactggat gcctgctacc 420  
tgaaggccct ggagggtctt gtcattggtg tcaccgcca gggagacatg gcttacctgt 480  
cggagaatgt cagcaaacac ctgggcctca gtcagctgga gctcattgga cacagcatct 540  
ttgatttcat ccaccctgt gaccaagagg agcttcagga cgccctgacc cccagcaga 600  
ccctgtccag gaggaagggt gagggcccca cggagcgggt cttctccttg cgcatgaaga 660  
gtacgctcac cagccgcggg cgacccctca acctcaaggc ggccacctgg aagggtgctga 720  
actgctctgg acatatgagg gcctacaagc cacctgcgca gacttctcca gctgggagcc 780  
ctgactcaga gcccccgctg cagtgcctgg tgctcatctg cgaagccatc cccacccag 840  
gcagcctgga gccccactg ggccgagggg ccttcctcag ccgccacagc ctggacatga 900  
agttcaccta ctgtgacgac aggattgcag aagtggctgg ctatagtccc gatgacctga 960  
tcggctgttc cgcctacgag tacatccacg cgctggactc cgatgcgggtc agcaagagca 1020  
tccacacctt gctgagcaag ggccaggcag taacagggca gtatcgcttc ctggcccgga 1080  
gtgtgtggta cctgtggacc cagaccagc ccacagtgtt gtcaggggga cggggccccc 1140  
agtgggagag tatcgtctgt gtccattttt taatcagcca ggtggaagag accggagtgg 1200  
tgctgtccct ggagcaaacg gagcaaacct ctgcgagacc cattcagcgg ggcgccccct 1260  
ctcagaagga caccctaacc cctggggaca gccttgacac ccctggcccc cggatccttg 1320  
ccttctgca cccgccttcc ctgagcgagg ctgcctggc cgctgacccc cgccgtttct 1380  
gcagccctga cctccgtcgc ctctgggac ccctcctgga tggggcttca gtagcagcca 1440  
ctcccagcac cccgctggcc acacggcacc cccaaagtcc tctttcggct gatctcccag 1500  
atgaactacc tgtgggcacc gagaatgtgc acagactctt caccctccgg aaagacactg 1560  
aggcagtgga gacagattta gatatagctc agatgaggaa actgaagctc agactgttga 1620  
ccacaggcac agaactcaga agtgatgggt ctgggacttc agccaaggtc cacccaagtc 1680  
caaggctcat cctcttacct ccctcctgcc ctccgcagga tgctgatgct ctggatttgg 1740  
agatgctggc cccctacatc tccatggatg atgacttcca gctcaacgcc agcgagcagc 1800  
taccagggc ctaccacaga cctctggggg ctgtcccccg gccccgtgct cggagcttcc 1860  
atggcctgtc acctccagcc cttgagccct ccctgctacc ccgctggggg agtgaccccc 1920  
ggctgagctg ctccagccct tccagagggg acccctcagc atcctctccc atggctgggg 1980  
ctcggaagag gaccctggcc cagagctcag aggacgagga cgagggagtg gagctgctgg 2040  
gagtgaagac tcccaaaagg tccccagcc cagaacacga aaactttctg ctctttcctc 2100  
tcagcctgag tttccttctg acaggaggac cagccccagg gagcctgcag gacccccactg 2160  
aacttaccba attccttctt tcagtcttaa gttttcccat tctagacccc taccctctag 2220  
gctgtgctgc tcctggactt catgcctctc cattctcatt gcctacaatc tctgtgcccc 2280  
agaacccct ccacttccca cccagccct ccagacatgc acttaccttg actttacccc 2340  
acatgttttg ggcacctggg gctccctcac cccttgggtg gtttgcaatc tgaagacttc 2400  
tccagccaca caggcacatg cacaggcacg gtgctgtctg catattgcca ggtggggaga 2460  
gaagccagga cccctcagct gtctgccacc atctatgtgc ctcccttacc cccagcttt 2520  
ctttctacag atggtgctac tcttgggtct ccacaggaaa aggcctcccc ccttcttagc 2580  
cccatttacc ccgtttgtgg aaggcactgc tcgctctgtt ttgtcagaga gtggcctatc 2640

```
cagattggtg ctatgggggg gtctgacccc tccctcctcc ctctggaggt gatgtggggc 2700
ctcaatggag ggaattgtgc tgggctaggg aaaggggagg gactagactg gccacactgg 2760
ctctgaaact caccaatctc tatacaccat aaagacctca ccttggtagg caccagaaaa 2820
aaaaaaaaa aaaaaaaatt                                     2840
```

CORRECTED VERSION

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
3 January 2003 (03.01.2003)

PCT

(10) International Publication Number  
WO 03/000864 A3(51) International Patent Classification<sup>7</sup>: C12P 21/06,  
C12N 1/20, 15/00, C07K 1/00, C07H 21/02

(21) International Application Number: PCT/US02/21179

(22) International Filing Date: 21 June 2002 (21.06.2002)

(25) Filing Language: English

(26) Publication Language: English

## (30) Priority Data:

60/300,518	22 June 2001 (22.06.2001)	US
60/301,787	29 June 2001 (29.06.2001)	US
60/301,792	29 June 2001 (29.06.2001)	US
60/301,892	29 June 2001 (29.06.2001)	US
60/301,893	29 June 2001 (29.06.2001)	US
60/303,405	6 July 2001 (06.07.2001)	US
60/303,442	6 July 2001 (06.07.2001)	US
60/364,438	15 March 2002 (15.03.2002)	US

(71) Applicant (*for all designated States except US*): INCYTE  
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo  
Alto, CA 94304 (US).

## (72) Inventors; and

(75) Inventors/Applicants (*for US only*): GANDHI, Ameena,  
R. [US/US]; 705 5th Avenue, San Francisco, CA 94118  
(US). SWARNAKAR, Anita [CA/US]; 8 Locksley Ave-  
nue #5D, San Francisco, CA 94122 (US). HAFALIA,  
April, J. A. [US/US]; 2227 Calle de Primavera, Santa  
Clara, CA 95054 (US). WARREN, Bridget, A. [US/US];  
2250 Homestead Court #2, Los Altos, CA 94024 (US).  
EMERLING, Brooke, M. [US/US]; 1735 Woodland  
Avenue #71, Palo Alto, CA 94303 (US). ARVIZU,  
Chandra, S. [US/US]; 1706 Morocco Drive, San Jose,  
CA 95125 (US). ISON, Craig, H. [US/US]; 1242 Weath-  
ersfield Way, San Jose, CA 95118 (US). HONCHELL,  
Cynthia, D. [US/US]; 400 Laurel Street #203, San Carlos,  
CA 94070 (US). LEE, Ernestine, A. [US/US]; 20523  
Crow Creek Road, Castro Valley, CA 94552 (US). YUE,  
Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087  
(US). FORSYTHE, Ian, J. [US/US]; 308 Roble Avenue,  
Redwood City, CA 94061 (US). RAMKUMAR, Jay-  
alaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA  
94555 (US). GRIFFIN, Jennifer, A. [US/US]; 33691  
Mello Way, Fremont, CA 94555 (US). YANG, Junming[CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US).  
SANJANWALA, Madhusudan, M. [US/US]; 210 Silvia  
Court, Los Altos, CA 94024 (US). BAUGHN, Mariah, R.  
[US/US]; 14244 Santiago Road, San Leandro, CA 94577  
(US). BOROWSKY, Mark, L. [US/US]; 122 Orchard Av-  
enue, Redwood City, CA 94061 (US). YAO, Monique, G.  
[US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US).  
CHAWLA, Narinder, K. [US/US]; 33 Union Square,  
#712, Union City, CA 94587 (US). BANDMAN, Olga  
[US/US]; 366 Anna Avenue, Mountain View, CA 94043  
(US). LAL, Preeti, G. [US/US]; P.O. Box 5142, Santa  
Clara, CA 95056 (US). BECHA, Shanya, D. [US/US];  
21062 Gary Drive #117, Castro Valley, CA 94546 (US).  
LEE, Soo, Yeun [KR/US]; 40 Westdale Avenue, Daly City,  
CA 94015 (US). RICHARDSON, Thomas, W. [US/US];  
616 Canyon Road #107, Redwood City, CA 94062 (US).  
ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way,  
San Jose, CA 95121 (US). LUO, Wen [CN/US]; 181  
Cymbidium Circle, South San Francisco, CA 94080 (US).  
TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose,  
CA 95118 (US). ZEBARJADIAN, Yeganeh [IR/US];  
830 Junipero Serra Boulevard, San Francisco, CA 94127  
(US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto,  
CA 94303 (US).(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,  
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,  
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,  
VN, YU, ZA, ZM, ZW.(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,  
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent  
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,  
NE, SN, TD, TG).

Published:

— with international search report

[Continued on next page]

(54) Title: NUCLEIC ACID-ASSOCIATED PROTEINS

(57) Abstract: Various embodiments of the invention provide human nucleic acid-associated proteins (NAAP) and polynucleotides which identify and encode NAAP. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of NAAP.

WO 03/000864 A3



**(88) Date of publication of the international search report:**

20 November 2003

**(48) Date of publication of this corrected version:**

18 December 2003

**(15) Information about Correction:**

see PCT Gazette No. 51/2003 of 18 December 2003, Section II

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
3 January 2003 (03.01.2003)

PCT

(10) International Publication Number  
**WO 03/000864 A3**(51) International Patent Classification<sup>7</sup>: **C12P 21/06**,  
C12N 1/20, 15/00, C07K 1/00, C07H 21/02

(21) International Application Number: PCT/US02/21179

(22) International Filing Date: 21 June 2002 (21.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/300,518	22 June 2001 (22.06.2001)	US
60/301,787	29 June 2001 (29.06.2001)	US
60/301,792	29 June 2001 (29.06.2001)	US
60/301,892	29 June 2001 (29.06.2001)	US
60/301,893	29 June 2001 (29.06.2001)	US
60/303,405	6 July 2001 (06.07.2001)	US
60/303,442	6 July 2001 (06.07.2001)	US
60/364,438	15 March 2002 (15.03.2002)	US

(71) Applicant (for all designated States except US): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **HAFALIA, April, J. A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **WARREN, Bridget, A.** [US/US]; 2250 Homestead Court #2, Los Altos, CA 94024 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue #71, Palo Alto, CA 94303 (US). **ARVIZU, Chandra, S.** [US/US]; 1706 Morocco Drive, San Jose, CA 95125 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **HONCHELL, Cynthia, D.** [US/US]; 400 Laurel Street #203, San Carlos, CA 94070 (US). **LEE, Ernestine, A.** [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **FORSYTHE, Ian, J.** [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **YANG, Junming** [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **SANJANWALA, Madhusudan, M.** [US/US]; 210 Silvia Court, Los Altos, CA 94024 (US). **BAUGHN, Mariah, R.**

[US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **CHAWLA, Narinder, K.** [US/US]; 33 Union Square, #712, Union City, CA 94587 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **LAL, Preeti, G.** [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **BECHA, Shanya, D.** [US/US]; 21062 Gary Drive #117, Castro Valley, CA 94546 (US). **LEE, Soo, Yeun** [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **LUO, Wen** [CN/US]; 181 Cymbidium Circle, South San Francisco, CA 94080 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **ZEBARJADIAN, Yeganeh** [IR/US]; 830 Junipero Serra Boulevard, San Francisco, CA 94127 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US).

(74) Agents: **HAMLET-COX, Diana et al.**; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:  
20 November 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NUCLEIC ACID-ASSOCIATED PROTEINS

(57) Abstract: Various embodiments of the invention provide human nucleic acid-associated proteins (NAAP) and polynucleotides which identify and encode NAAP. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of NAAP.



WO 03/000864 A3

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/21179

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12P 21/06; C12N 1/20, 15/00; C07K 1/00; C07H 21/02  
US CL : 435/69.1, 252.3, 320.1; 530/350; 536/23.1

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 252.3, 320.1; 530/350; 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
STN AND WEST. Sequence data bases search for SEQ ID Nos. 1 and 37.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X — A	Database on Genbank, SINGH et al. Identification of 10 murine homeobox genes, 01 July 1993, Accession No. P31316. Accession No. P31316 is 87% identical to Applicants SEQ ID NO : 1, reads on fragments of claim 1(h) and (i) and claim 17.	1, 17 ----- 2-10, 12-13, 18, 56, 92
X — A	Database on Genebank, EST, 10 August 2000, Accession No. BE551998. Accession No. BE551998 is 48.1% identical to Applicants SEQ ID NO : 37, having 60 contiguous nucleotides reads upon claim 3 & 13.	3, 13 ----- 1-10, 12, 17-18, 56, 92

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

19 March 2003 (19.03.2003)

Date of mailing of the international search report

18 AUG 2003

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Valerie Bell-Harris  
Tekchand Saidha

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/21179

## Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:  
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-10, 12-13, 17-18 & 92 (all partly)

Remark on Protest

☐  
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1)) (July 1998)

**BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING**

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group 1, claim(s) 1-10, 12-13, 17-18, 56 & 92, drawn to polypeptide of SEQ ID NO: 1, encoding DNA of SEQ ID NO : 37, host cells and method of making the polypeptide.

Groups 2-36, claim(s) 1-10, 12-13, 17-18, 57-91 & 93-127, drawn to polypeptide of SEQ ID NOS : 2-36, encoding DNA of SEQ ID NO : 38-72, host cells and method of making the polypeptide, respectively. [each of polypeptide sequences of SEQ ID NOS : 2-36 and the corresponding DNA sequences of SEQ ID NOS : 38-72 form the 35 groups hich have been put together for convenience]

Group 37, claim(s) 11, 30-45, drawn to antibody against SEQ ID NO : 1.

Groups 38-72, each group corresponding to claim(s) 11, 30-45, drawn to antibody against each of the sequences of SEQ ID NOS : 2-36.

Group 73, claim(s) 14-16, drawn to method of detection of polymucleotide of SEQ ID NO : 37.

Groups 74-108, claim(s) 14-16, drawn to method of detection of polymucleotide for each of the sequences of SEQ ID NOS : 38-72, respectively.

Group 109, claim(s) 19, drawn to method of treatment using polypeptide of SEQ ID NO : 1.

Groups 110-144, claim(s) 19, drawn to method of treatment using polypeptides of each of the sequences of SEQ ID NOS : 2-36, respectively.

Group 145, claim(s) 20-25 & 27, drawn to a method of screening agonist/antagonist of the polypeptide of SEQ ID NO : 1.

Groups 146-180, claim(s) 20-25 & 27, drawn to a method of screening agonist/antagonist of the polypeptide of each of the sequences of SEQ ID NOS : 2-36, respectively.

Group 181, claim(s) 26, drawn to a method of screening a compound that specifically binds to polypeptide of SEQ ID NO : 1.

Groups 182-216, claim(s) 26, drawn to a method of screening a compound that specifically binds to each of the polypeptide of SEQ ID NOS : 2-36, respectively.

Group 217, claim(s) 28-29, drawn to a method of screening a compound for assessing toxicity or effectiveness in altering the target nucleotide sequence of SEQ ID NO : 37.

Groups 218-252, claim(s) 28-29, drawn to a method of screening a compound for assessing toxicity or effectiveness in altering the target nucleotide for each of the sequences of SEQ ID NOS : 38-72, respectively.

Group 253, claim(s) 46-55, drawn to a microarray and method of generating expression profile using nucleotide sequence of SEQ ID NO : 37.

Group 254-288, claim(s) 46-55, drawn to a microarray and method of generating expression profile using nucleotide sequences of SEQ ID NOS : 38-72, respectively. The inventions listed as Groups 1-288 do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group 1 has the special technical feature of the nucleotide sequence of SEQ ID NO : 37 and the encoded polypeptide of SEQ ID NO : 1, host cell and a method of making the polypeptide, which Groups 2-288 do not share. For the same reason, Group 2-36 has the special technical feature of the nucleotide sequences of SEQ ID NOS : 37-72 and the encoded polypeptides of SEQ ID NOS : 2-

## INTERNATIONAL SEARCH REPORT

PCT/US02/21179

36, respectively, host cell and a method of making the polypeptide, which Groups 37-288 do not share. Each of the Groups 37-72 have a special technical in a distinct antibody pertaining to each of the polypeptide sequences of SEQ ID NOS : 1-36, which Groups 73-288 do not share. Each of the Groups 73-288, employ structurally distinct nucleotide sequences of SEQ ID NOS : 37-72 or the polypeptide sequences of SEQ ID NOS : 1-36 in distinct methods, however, in view of 37 CFR 1.475(b), when claims corresponding to different categories of inventions are present then only (3) and additional methods of use are deemed to lack unity. Thus the various groups discussed above show a lack of unity of invention.

**THIS PAGE BLANK (USPTO)**